DOCUMENT RESUME

ED 253 584 TM 850 130

AUTHOR	Clark, Mary Jo; Grandy, Jerilee
TITLE	Sex Differences in the Academic Performance of Scholastic Aptitude Test Takers.
SPONS AGENCY	College Entrance Examination Board, New York, N.Y.
REPORT NO	CB-R-84-8; ETS-RR-84-43
PUB DATE	84
NOTE	37p.
AVAILABLE FROM	College Board Publications, Box 886, New York, NY 10101 (Report No. 84-8, $5.00).
PUB TYPE	Information Analyses (070) -- Reports - Research/Technical (143)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Academic Achievement; Cognitive Ability; College Bound Students; *College Entrance Examinations; Educational Trends; Females; Higher Education; High Schools; Intellectual Disciplines; *Performance Factors; *Predictive Validity; Scores; *Sex Differences; Test Bias
IDENTIFIERS	*Scholastic Aptitude Test

ABSTRACT

The average College Board Scholastic Aptitude Test (SAT) scores for women have declined more than the scores for men. Recent evidence concerning the academic performance of men and women was studied by examining sex differences among all SAT takers, test takers grouped by anticipated major field of study, and college freshman year courses and grades. The research was conducted to determine whether there are consistent differences in the intellectual abilities of men and women; whether precollege admissions variables predict college performance with equal accuracy for both sexes; and whether the content or structure of the SAT contributes to observed sex differences in test performance. Neither differences in high school courses and grade point average, major field and career interests, and socioeconomic backgrounds, nor research evidence on differential cognitive functioning, were found sufficient to account for all of the observed sex differences in SAT performance. This study concludes that sex-related SAT differences are very small relative to the generally similar levels of performance by men and women, and that using test scores and high school records to predict first-year college grades continues to be reasonably effective for both sexes. (Author/DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED253584

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

S. W. Gay Ames

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

College Board Report

No. 84-8

Sex Differences in the Academic Performance of Scholastic Aptitude Test Takers

Mary Jo Clark
Jerilee Grandy

Sex Differences in the Academic Performance of Scholastic Aptitude Test Takers

Mary Jo Clark
Jerilee Grandy

College Board Report No. 84-8

ETS RR No. 84-43

College Entrance Examination Board, New York, 1984

Mary Jo Clark is currently Director of Program Review and Development, New Mexico Highlands University, Las Vegas, New Mexico. Previously she was Research Psychologist, Educational Testing Service, Princeton, New Jersey.

Jerilee Grandy is Associate Research Scientist in the Division of Measurement Research and Services, Educational Testing Service, Princeton, New Jersey

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

The College Board is a nonprofit membership organization that provides tests and other educational services for students, schools, and colleges. The membership is composed of more than 2,500 colleges, schools, school systems, and education associations. Representatives of the members serve on the Board of Trustees and advisory councils and committees that consider the programs of the College Board and participate in the determination of its policies and activities.

Additional copies of this report may be obtained from College Board Publications, Box 886, New York, New York 10101. The price is $5.

Copyright © 1984 by College Entrance Examination Board. All rights reserved.
The "College Board" and the acorn logo are registered trademarks of the College Entrance Examination Board.
Printed in the United States of America.

CONTENTS

Abstract 1
Introduction 1
Sex Differences among SAT Takers 2
Literature on Cognitive Functioning 3
Differential Performance on the SAT 4
Changes in the SAT Examinee Population 4
Study of Two SAT Subpopulations 7
Predictive Validity 18
Correlations with First-Year Grades 19
Under- and Overprediction 20
Contents and Structure of the SAT 22
Verbal Performance 22
Mathematical Performance 23
Test Construction and Administration 24
Summary and Conclusions 25
References 26

Figures

1. SAT-Verbal Score Distributions 5
2. SAT-Mathematical Score Distributions 7
3. Trends in Eight Ability Variables for Social Science Majors 1978–1982 12
4. Trends in Six Background Variables for Social Science Majors 1978–1982 13
5. Trends in Eight Ability Variables for Mathematics Majors 1978–1982 16
6. Trends in Six Background Variables for Mathematics Majors 1978–1982 17

Tables

1. Comparisons of Numbers of High and Low Scoring SAT Examinees Between 1973 and 1982 5
2. Samples Selected from SRS Files 8
3. Statistics on Students Planning to Major in Political Science or Sociology, 1978–1982 ($N \cong 5{,}300$) 10
4. Statistics on Students Planning to Major in Mathematics or Statistics, 1978–1982 ($N \cong 6{,}600$) ... 14
5. Characteristics of Universities and Studies in Predictive Validity Analyses 19
6. Multiple Correlations Between Predictors and First-Year College Grades in 41 Validity Studies of 1980 Freshman Students 20
7. Accuracy of First-Year College Grade Predictions for Groups of Men and Women in the 1980 College Board Validity Study Service 21
8. Comparative Averages for Males and Females on the Mathematical Sections of the SAT by Years of Preparation (Based on SDQ Results, 1978–79) 24

ABSTRACT

The number of women college students has increased dramatically over the past fifteen years; in this same time period, the average College Board Scholastic Aptitude Test (SAT) scores for women have declined more than the scores for men. This study summarizes recent evidence concerning the academic performance of women and men by examining sex differences among (1) all SAT takers, (2) test takers grouped by anticipated major field of study, and (3) college freshman year courses and grades.

Three major research questions guided this study: Are there consistent differences in the intellectual abilities of men and women? Do precollege admissions variables predict college performance with equal accuracy for women and men? Have the contents or structure of the SAT contributed to observed sex differences in performance on the test?

Though the men and women who take the SAT differ in high school courses and grade point average, major field and career interests, and socioeconomic backgrounds, neither these differences nor research evidence on differential cognitive functioning were found sufficient to account for all of the observed sex differences in performance on the SAT. Some but not all of the male–female differences in scores on the mathematics portion of the SAT were accounted for by differences in the depth and recency of course preparation. The observed first-year college grades of women were slightly underpredicted by test scores and high school grades, but the magnitude of the underprediction was very small. While specific changes in the contents or structure of the SAT could not be associated with score changes, it seems likely that numerous small adjustments have played some part in recent shifts among the average test scores of women and men.

Consistent with recent literature on sex differences in cognitive performance, this study concludes that sex-related SAT differences are very small relative to the generally similar levels of performance by men and women, and that using both test scores and high school records to predict first-year college grades continues to work reasonably well for both sexes.

INTRODUCTION

The last 15 years have seen shifts in the college attendance and career directions of both high school age and older women in the United States, associated largely with changing social attitudes and self-conceptions of the academic and career pursuits that are appropriate and available to them. For the most part, these shifts have been supported and encouraged by U.S. high schools and colleges; for example, studies have been carried out and policies revised to reduce sex bias in career counseling, to encourage women to continue the study of mathematics in high school, and to reduce admission and attendance barriers at the college level.

Some of these shifts are reflected in the increasing numbers of women who have taken the College Board's Scholastic Aptitude Test (SAT) over the past several years, as indicated in the annual issues of the Admissions Testing Program (ATP) *National Report*. Comparing 1972 with 1981, the number of women in the ATP increased from about 527,000 to about 545,000; during the same period of time, the number of men declined from almost 555,000 to about 505,000 (*National College-Bound Seniors* 1981, p. 4). These shifts in ATP participation have been accompanied by changes in the overall average SAT scores of men and women. Comparing test performance in 1972 with 1983, we find declines in the average SAT-verbal scores from 454 to 430 (24 points) for men and from 452 to 420 (32 points) for women; the decline in average SAT-mathematical scores since 1972 also were greater for women, from 461 to 445 (16 points), than for men (505 to 493, 12 points). Yet both the median percentile high school rank and the mean high school grade point average (GPA) have remained higher for women than for men, although there was an overall decline in the high school ranks of college-bound seniors during this period. Looking back even further in the history of SAT performance, we find that women averaged SAT-verbal scores slightly higher than the verbal scores of men until 1972, when the average verbal score of college-bound males was higher than the average verbal score of college-bound females for the first time. After 1972, the difference increased to a 10-point advantage for males by 1983. Males have consistently outperformed females on the mathematical test by 40 to 50 points in the period 1967 to 1983 (*National College-Bound Seniors* 1983, pp. 4–6).

Though the College Board routinely reports changes in the characteristics of the ATP population and shifts in performance like the ones previously noted, no systematic investigation in recent years has been made concerning the reasons for the changes or the implications of the patterns of test scores and high school grades for college admission and performance. This report reviews the available evidence on sex differences in academic performance and evaluates the results in relation to recent literature on sex-related differences in cognitive functioning as well as the literature on possible bias in the use of tests to predict college performance. The first section provides background information on SAT score changes and the characteristics of the men and women who take the test. The second section briefly reviews the evidence for gender-differentiated cognitive functioning. Are there consistent and dependable differences in the intellectual abilities of men and women? In the third section, the shifting composition of cohorts of test takers, changing interests in course elections and major fields, and other descriptive data supplied by test registrants are examined in an effort to help explain the observed differences in performance by sex. The fourth section

examines the academic performance of men and women in the first year of college in relation to precollege test scores and the high school record. Do precollege admissions variables predict college performance with equal accuracy for women and men? Finally, in the fifth section, the contents and structure of the SAT-verbal and SAT-mathematical sections are examined for possible clues regarding the score differences of men and women. Changes in item types, item wording, the contents of reading passages, and equating procedures are made periodically. Have subtle changes in the test contributed to the greater decline of test scores among women than men? Do the current test contents contribute to higher average scores for men on both the verbal and mathematics tests?

The next section reviews the current status of sex differences in the academic performance of those men and women who take the SAT.

SEX DIFFERENCES AMONG SAT TAKERS

The overall decline in SAT scores received major attention in the mid-1970s from a blue-ribbon panel appointed by the College Board to study the causes and impact of the changing scores. In their final report (*On Further Examination* 1977), the panel reviewed possibilities ranging from psychometric technicalities to changes in the composition of the student group taking the test to social changes such as the pervasiveness of television. In general, the panel concluded that, if anything, the psychometric "scale drift" of the verbal and mathematical tests minimized rather than accounted for score changes in the 1960s and 1970s; that changes reflecting increasing numbers of poorly prepared test takers occurred mostly in the 1960s, with "compositional" shifts slowing materially by the early 1970s; and that the continuing across-the-board score declines apparently were related to more "pervasive" changes in the schools and in society. The disparity between the mathematical scores of men and women were considered to reflect "more than anything else the traditional sex stereotyping of career opportunities and expectations" (p. 16). No mention was made of the shift from female to male higher average verbal scores.

A report on declining ACT scores for use in college admissions (Munday 1976) also found marked sex differences, with the proportion of women test takers increasing from 45 percent in 1964–65 to 52 percent in 1973–74. During this time, the ACT composite mean score for men declined from 20.2 to 19.3 while the composite mean score for women declined from 19.7 to 18.1. What had been a 0.5 score advantage for men in 1964–65 increased to a 1.2 score advantage by 1973–74, as the average scores of women slid more rapidly than those of men. Nevertheless, women continued to average almost 2 points higher on the English section of the test, while men consistently averaged more than 2 points higher on the mathematical test. This and other studies confirmed that the score decline was not just a phenomenon of the SAT.

Since 1972, more than 80 percent of SAT takers have filled out the Student Descriptive Questionnaire (SDQ), which covers some 70 informational items and provides a relatively complete picture of the test-taking population each year. The SDQ data report only on persons taking the test; therefore, shifts from year to year reflect shifts in self-selection patterns among test takers. Despite this limitation, the data have been invaluable in associating the backgrounds and interests of students with patterns of scores on the College Board battery of aptitude and achievement tests. Justice to the richness of these comparisons is beyond the scope of this report; detailed descriptions can be found in College Board publications such as *National College-Bound Seniors* (1983) and *Profiles, College-Bound Seniors* (Ramist and Arbeiter 1984).

In addition to the average test score differences reported above, some of the highlights of differences between male and female ATP registrants in recent years, as summarized by Ramist and Arbeiter (1984, pp. v–vi), include:

- The number of female registrants increased from 48.9 percent in 1972 to 52.0 percent in 1981, declining slightly to 51.8 percent in 1983.
- A language other than English is more likely to be the best language for males (4.4 percent) than for females (3.5 percent).
- Males are slightly more likely to attend nonpublic schools (21 percent compared to 19 percent) and to be in an academic or college preparatory program (78 percent compared to 76 percent).
- Males take more years of study in academic subjects (a total of 16.48 years) than females (16.18 years), though large increases by females in physical sciences and mathematics courses since 1980 is reducing this difference.
- Females have a higher mean GPA, 3.11, and median high school rank, 75.8, than males (3.00 and 74.2, respectively), and are more likely to receive a high school honor or award (62 to 55 percent).
- Males are more interested in graduate study than females (46 percent compared to 42 percent). However, there has been a large increase in female interest in graduate study, from 30 percent in 1973 to 42 percent in 1983.
- Choices of intended area of study differ greatly by sex. For males, the most popular choice is engineering (22 percent, but only 4 percent of females) followed by business and commerce (17 percent), computer sciences (12 percent), health and medical (9 percent), and social sciences (7 percent). For

females, the most popular choices are health and medical (20 percent), business and commerce (20 percent, up from 7 percent in 1973), computer sciences (9 percent), social sciences (7 percent), and education (7 percent).

It should be noted that the SDQ data are self-reported by test registrants; although numerous studies have found moderate to high correlations (.70 to .90) between self-reports and corresponding objective measures, recently Hamilton (1981) reported systematic sex differences in self-report errors among college students when they were asked for their test scores and GPA. Validity coefficients for both sexes were reasonably high (median of .79), but women tended to inflate their scores more than men and to make errors that were more dispersed. Hamilton concluded that, if this finding is true for SDQ data generally, using the data to examine substantive sex differences or similarities will introduce nonrandom errors with complex statistical consequences. Since many of the data presented in the following sections of this report originated in SDQ reports, this possible confounding effect should be kept in mind when evaluating the results.

One of the original plans for this report was to evaluate the impact of increasing numbers of older college students on average SAT scores and the usefulness of the scores in predicting college grades for older students. However, this did not prove feasible for a number of reasons. First, almost all of the SAT test takers are high school students or very recent graduates; very few adults take the test (less than 2 to 3 percent of the test takers in recent years have been over 20 years old). Second, almost all of the College Board reports on scores and test-taker characteristics focus on "college-bound seniors" in order to minimize test repeaters in the file and to maximize the usefulness of the information in relation to college admission. Since most adults who take the test are well beyond high school senior status, they are omitted from the tabulations. The College Board has sponsored one recent study of the academic performance of older students (Casserly 1982) which looked at the predictive validity of test scores by age group and sex at three universities that enrolled relatively large numbers of older freshmen. Multiple correlations of test scores and high school record with first-year grades averaged a high of .56 for younger females to a low of .45 for older females. Casserly concluded that among the colleges she studied, test scores were useful predictors of college performance for older applicants.

The best evidence of predictive accuracy for college freshmen of different ages comes from data in the ACT validity service files of 1977 (Levitz 1981). This study examined actual and predicted grades for 107,059 freshmen at 223 institutions in age groups 17–19, 20–25, and 26 and older. The median cross-validated correlation coefficient measuring the relationship between predicted and actual GPA was .57 for the 17–19 age group, .32 for freshmen age 20–25, and .44 for those age 26 and older. Age alone explained only 4 percent of the variance in the ACT composite score in data drawn from both 1972 and 1977. Unfortunately, the predictive validity data were not analyzed separately by sex.

Just as the panel on score decline found many possible reasons for declining SAT scores among high school seniors, there are many possible reasons for systematic differences in the scores of women and men. The next section examines the evidence for the possibility that the test scores reflect basic differences in the cognitive abilities of the sexes.

LITERATURE ON COGNITIVE FUNCTIONING

The "bible" on gender differences is the review by Maccoby and Jacklin (1974) titled *The Psychology of Sex Differences*. Though published 10 years ago, it is still the most complete and accurate source of information on the topic. Maccoby and Jacklin concluded that three cognitive gender differences were "well established" in the literature: Females exhibit greater verbal ability than males, and males have better mathematical ability and visual-spatial ability than females. In general, the differences increase substantially with the onset of adolescence. A number of more recent reviews support these conclusions, though they tend to emphasize the small size of the gender differences and the likely contributing influence of situational factors. These more recent reviews are summarized briefly in this section.

Hyde (1981) used meta-analytic procedures to review the studies cited in Maccoby and Jacklin in order to assess the magnitude of gender differences using statistical indexes of effect size. She found that gender differences accounted for only about 1 percent of the variance on verbal and mathematical ability, and about 4 percent of the variance in visual-spatial ability. Hyde concluded that differences of this small size have practical implications only in two areas: vocational counseling, where the temptation to use "well-established" gender differences may result in poor advice; and the explanation of observed sex differences in certain highly selective occupations, where relatively small mean differences can generate rather large differences at the tails of distributions. In Hyde's words:

> In sum, even small average gender difference in abilities can generate rather large differences in proportions of males and females above some high cutoff point that might be required for outstanding performance of some occupation. Nonetheless, the known differences in abilities are still too small to explain the observed occupational differences (p. 899).

Recently, Wittig and Petersen (1979) examined developmental issues concerning sex-related differences in cognitive functioning. Their review summarized what we know about the possible influences of genes, brain organization, hormones, socialization, and measurement on the development and performance of cognitive skills. Each

of these categories of analysis was found to shed light on the problem; no one of them offered a satisfactory explanation for the small but consistent sex differences that were found. A number of methodological problems in research on sex differences were identified and discussed, including questionable sampling procedures in many studies; development versus chronological age comparisons; varying definitions of skills to be measured; and statistical versus practical differences. A final chapter by Jacklin attempted to summarize the critical conclusions of the different authors and to place the research in the context of its social significance. The authors conclude that more research is needed before we have a clear and accurate picture of how sex-related differences in cognitive functioning may develop.

The importance of experience as a mediating variable in the development of cognitive skills has been suggested by a number of researchers as a possible explanation for the inconsistency of findings across studies of sex-related differences in cognitive functioning, and for sex differences in responsiveness to special training efforts (M.C. Linn et al. 1981). There are also psychosocial differences between the sexes in areas such as confidence, anxiety, and the attribution of causality about success and failure that might help to explain differential academic achievement, and there are different expectations of males and females by families, schools, and society. All of these factors contribute to the complexity of the problem without shedding much light on it.

Deaux (1984) noted that the literature related to sex or gender has proliferated in the past decade, and that it can generally be classified into research on (1) sex as a subject variable; (2) individual differences in masculinity, femininity, and androgyny; and (3) sex as a social category. In reviewing differences between men and women, the author concluded that:

> Sex of subject differences are less pervasive than many have thought. Main effects of sex are frequently qualified by situational interactions, and the selection of tasks plays a critical role in eliciting or suppressing differences. Furthermore, the amount of variance accounted for by sex, even when main effects are reliable, is typically quite small. Thus, when any particular behavior is considered, differences between males and females may be of relatively little consequence. Approaching sex as a subject variable, although a popular pursuit with a long history, may not be the most productive route for understanding gender, at least in its simplest form (p. 110).

Deaux argued that more attention should be directed at the choices men and women make, as opposed to the capabilities that they show in a particular domain. As she points out, "when one looks for differences between women and men engaged in a specific activity, relatively few are found" (p. 114).

To summarize, the research literature either finds no difference between men and women in performance on cognitive skills, or finds a slight advantage for females on verbal skills and a slight advantage for males on mathematical and spatial skills. Male-female biological differences do not appear to explain the observed differences in cognitive functioning; experiences, stereotypes, and expectations no doubt play a role, but it has been difficult to identify specific ways in which they may account for differences in academic performance. In addition, the measures we use may contribute to the differences we observe. This last possibility is considered in greater detail in the section "Contents and Structure of the SAT."

This brief review of the research literature on sex differences in cognitive functioning provides a context for the next section in which we examine the academic performance of college-bound SAT takers in greater detail.

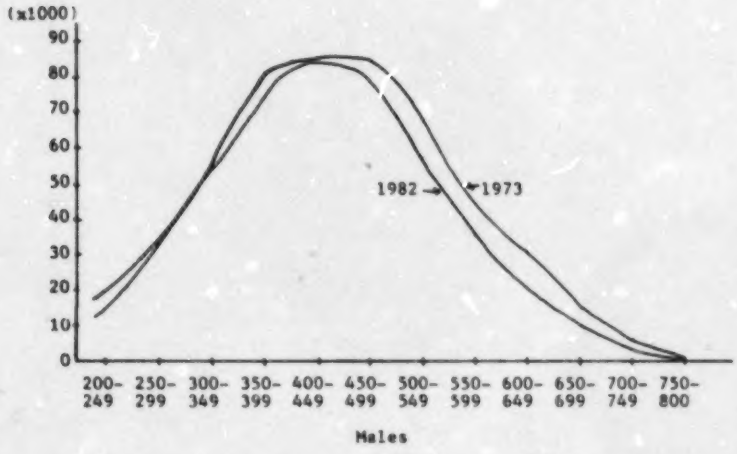
DIFFERENTIAL PERFORMANCE ON THE SAT

One suggestion that arises repeatedly in the literature on score decline and in the literature on sex differences is that test performance is population dependent. The score decline may indicate a shift of some sort in population characteristics. We have attempted to address the issue of population dependency by first examining shifts in SAT examinee populations from 1973 to 1982 and then by examining two narrowly defined subpopulations to see if there are sex differences in test performance in those groups.

Changes in the SAT Examinee Population

One question that arises in most discussions of score decline is whether there has been an overall decline in the academic ability of college students or whether the score decline has been brought about by an increasingly large number of low-ability applicants. In the first case, we would expect the entire distribution of scores to shift to the left. The number of high-scoring students would decline by about the same amount as low-scoring students would rise. In the second case, there would be a change in the shape of the distribution with an increase in the number of low-scoring students and only a small decrease in the number of high-scoring and middle-scoring students.

Figure 1 shows graphically how the distributions of SAT-verbal scores have changed for males and females separately. For males there has been an increase in the number of low-scoring students and a decline in the number of students scoring in the mid-range and higher. Part of this decrease is accounted for by the decrease in total candidate volume. If we visually shift the 1982 curve up slightly, we have a more accurate picture of the score ranges showing a relative decrease in students. It appears that there are relatively fewer males scoring in the entire range above the median. This, in fact, is confirmed by the actual percentages in each score range.



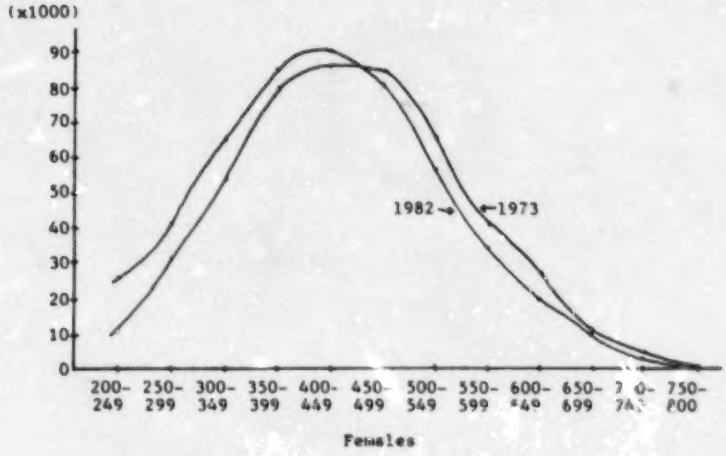


Figure 1. SAT-Verbal Score Distributions.

Table 1. Comparisons of Numbers of High and Low Scoring SAT Examinees Between 1973 and 1982 (Numbers in Parentheses are Percents of the Total Population)

A. SAT-Verbal

Numbers of Males				*Numbers of Females*			
Score	*1973*	*1982*	*Change*	*Score*	*1973*	*1982*	*Change*
≥600	52,859 (10)	36,760 (8)	−16,099	≥600	45,397 (9)	33,688 (7)	−11,709
300	42,615 (8)	52,845 (11)	+10,230	300	41,297 (8)	67,009 (13)	+25,712

B. SAT-Mathematical

Numbers of Males				*Numbers of Females*			
Score	*1973*	*1982*	*Change*	*Score*	*1973*	*1982*	*Change*
≥600	115,451 (22)	101,966 (21)	−13,485	≥600	53,578 (11)	48,856 (10)	− 4,722
350	52,343 (10)	58,272 (12)	+ 5,929	350	77,206 (15)	105,789 (21)	+28,583

Table 1 shows the actual number of males scoring in the highest range (600 and above) compared to the number in the lowest range (under 300). From 1973 to 1982, the percentage in the high range dropped from 10 to 8 percent of all males taking the SAT-verbal sections. Over the same period, the number scoring under 300 rose from 8 to 11 percent of that population. It appears, therefore, that for males there has been a shift to the left in the distribution with the total male population scoring lower than it did 10 years ago.

For females there has been both a shift to the left in the distribution and a large increase in the number of low-scoring students. All score ranges below the median are showing larger numbers of students, while score ranges above the median are showing fewer. From Table 1 we see that the number scoring over 600 has declined from 9 to 7 percent of the female population, but the number scoring under 300 has risen from 8 to 13 percent. This is a considerable rise in the number of very low-scoring students.

In 1973 the mean SAT-mathematical score was 502 for males and 460 for females. By 1982 it was 493 for males and 443 for females. The apparent decline in mathematics ability was therefore 8 percent of a standard deviation for males and 16 percent of a standard deviation for females. Figure 2 shows the distributions of SAT-mathematical scores for each sex. The shapes of these curves are quite irregular; the 1982 curve for males shows fewer students scoring in the 500–549 range than in the ranges just above and below. This suggests a possible bimodal distribution. The curve for females also shows a dip in that range. This was not the case in 1973 for either sex, and we cannot account for the shapes of the 1982 distributions from the data at hand.

Males show only a small change in the percentage scoring over 600 on the SAT-mathematical sections (Table 1). In 1973, 22 percent of all males scored 600 or higher; in 1982, 21 percent still scored in this range. The number scoring under 350 has risen from 10 to 12 percent of the male population. The greatest difference is in the midrange, between 450 and 600. The decrease in candidate volume is primarily in this range.

It is clear from Figure 2 that among females, there has been a very great increase in the number of examinees scoring under 400 and an even larger decrease in the number scoring between 500 and 600. This distribution shows both a shift to the left and an increase in the number of low-scoring students. In 1973, 11 percent of the female population scored over 600 on the SAT-mathematical sections (Table 1). By 1982, that percentage had only dropped to 10 percent. But the number scoring under 350 had risen from 15 to 21 percent of all females taking the exam. It is only in the range above 600 that females have nearly maintained their numbers.

High school GPA and class rank have been consistently higher for females than for males. For males, the average GPA rose from 2.95 in 1973 to 3.05 in 1976 and 1977. It then declined to 3.00 in 1980 and has remained at that level through 1982. For females, the average GPA rose from 3.13 to 3.18 then declined to 3.11. Class rank showed a similiar pattern for males, but it was erratic for females. For males it averaged the 75.6 percentile in 1973 and the 74.2 percentile in 1982. The mean percentiles for females were 80.8 in 1973 and 75.5 in 1982. Both sexes showed a decline in class rank, therefore, with females showing the greater decline. In GPA, however, males showed an overall rise. It is difficult to know how to interpret the patterns in GPA and class rank because they are so erratic. Possibly they were confounded by grade inflation and other factors. After all, there is no attempt to equate grades from year to year as we equate test scores. It is quite likely that a "B" one year does not correspond exactly to a "B" the next year.

Scores on the Test of Standard Written English (TSWE) were consistently higher for females than for males. This would seem to contradict the findings for SAT-verbal sections, since both are measures of some aspects of verbal ability. Average scores on the TSWE declined from 422 to 417 for males between 1975 and 1982. For females they declined from 443 to 428 during this period. Although the decline was greater for females than for males (15 points versus 5 points), females still had a mean that was 11 points higher than the mean for males.

Females took fewer mathematics courses, on the average, than did males, though the difference between them became smaller.

More males than females aspired to graduate or obtain professional degrees. In 1973, 23 percent of the males and only 11 percent of the females indicated that they planned to obtain a doctorate or professional degree. In 1982 this figure had dropped to 18.9 percent for males and had risen to 16.4 percent for females. The number who were undecided about their degree goals had dropped about equally for both sexes, from 27 to about 19 percent.

Major fields of interest also changed over the past decade. In 1973, 13 percent of the males and 7 percent of the females planned to major in business. By 1982, these figures had risen to 17.5 and 19.8 percent, respectively. Business became the single most popular major field for females, whereas in 1973 it was nursing. Among males, engineering rose in popularity from 11 percent in 1973 to 22.5 percent in 1982. Engineering became the most common major for males.

There was a change in the racial/ethnic composition of the examinee population, and the proportions were slightly different for each sex. In 1973, 11 percent of the males and 14 percent of the females identified themselves as minorities. In 1982, these figures were 17.4 and 19.1 percent, respectively. The percentage who were black rose from 6 to 7.5 percent for males and from 9 to 10.1 percent for females.

What all of these statistics suggest rather conclusively

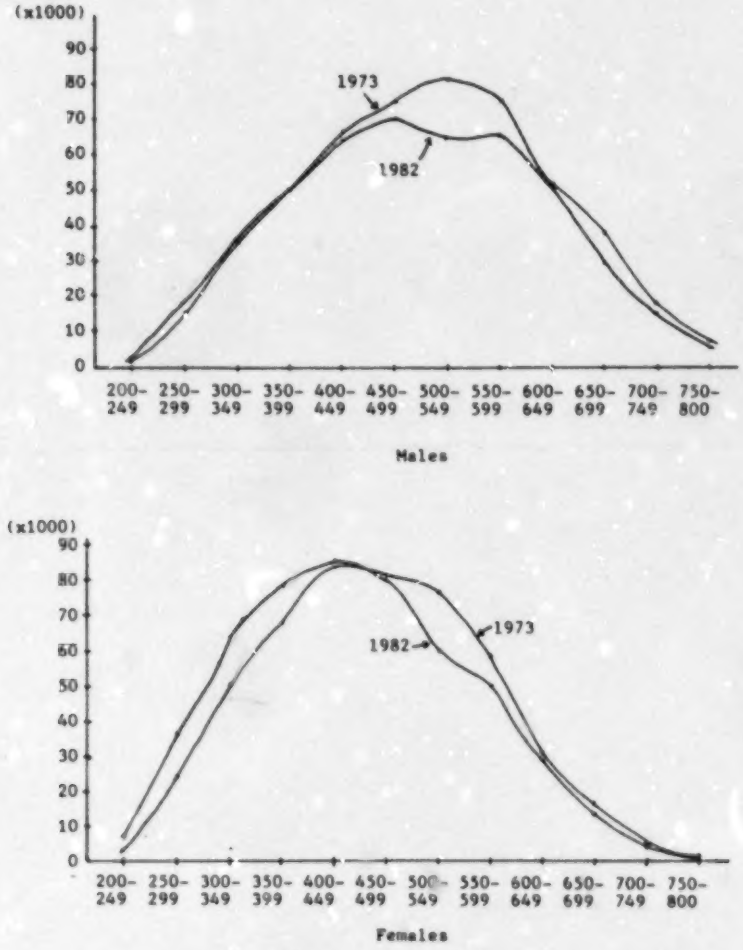


Figure 2. SAT-Mathematical Score Distributions.

is that the population of SAT examinees changed in very real ways. Its ethnic composition changed noticeably, students' interests changed, their aspirations changed, and their test scores changed. Not only did all of these student characteristics change, but they changed differently for males and females. If we are to gain any understanding of the real meaning of the score decline and of the differences in the scores of males and females, we will need to study relatively homogeneous subgroups of the examinee population to see if or how those subgroups changed over time.

Study of Two SAT Subpopulations

Purpose

By examining trends in SAT scores and other variables over a 10-year period, we saw that the population of students taking the SAT has changed in many ways and that males and females differ with respect to these variables. From looking at those data alone, we could not account for the score decline or for the observed sex differences. What was needed was a study of a definable subgroup of the population, a group that was homogeneous in some respect and could be compared with a similar group from year to year. Because students vary so greatly in interests, we selected our groups on the basis of major field of interest. If we found that there was a score decline among mathematics majors, or that male mathematics majors had much higher SAT-mathematical scores than female mathematics majors, we would be greatly concerned because of what these statistics might imply about future mathematicians. We could then look at background characteristics to see what variables might account for the score trends, and what experiences might account for observed sex differences.

For this study we selected two samples that we would

Table 2. Samples Selected from SRS Files*

Year	Number of Math Majors on File	Number of Social Science Majors on File	Total Number of Math Majors Sampled	Total Number of Social Science Majors Sampled	Number of Math Majors with SAT Scores				Number of Social Science Majors with SAT Scores			
					White Male	White Female	Non-White Male	Non-White Female	White Male	White Female	Non-White Male	Non-White Female
1978	13,664	10,511	1,705	1,301	780	636	119	96	475	504	105	144
1979	11,414	8,643	1,422	1,076	661	492	111	104	358	420	102	130
1980	10,315	8,126	1,287	1,015	532	514	74	92	375	373	82	116
1981	9,894	8,337	1,235	1,041	506	517	82	67	390	398	76	106
1982	9,567	8,104	1,195	1,012	511	487	67	85	397	376	62	113

*Math majors are defined as those answering SDQ item 63 as mathematics or statistics (codes 625 or 626). Social science majors are defined as those specifying either political science or sociology (codes 785 or 789).

expect to have very different interests and experiences. We required that our samples be from two traditional areas, one primarily requiring verbal skills and the other requiring quantitative skills. Furthermore, we required samples in which the numbers of male and female examinees were about equal, and in which the numbers were neither greatly rising nor declining over time. The groups that best satisfied these conditions were (1) political science and sociology majors and (2) mathematics and statistics majors. They were identified by their responses to item 63 in the Student Descriptive Questionnaire (SDQ). For simplicity, we will refer to the first group as social science majors and to the second group as mathematics majors.

Sampling

SAT scores along with SDQ responses are retained by the College Board on computer files referred to as the Summary Reporting System (SRS). At the time we undertook this study, data collected prior to 1978 were not available for analysis. We selected our samples, therefore, from 1978 to 1982 files. This was done by selecting every eighth record from among all those with the selected major field interests. Table 2 summarizes the characteristics of each sample selected and the size of its sampling frame for each year.

Analyses

For each of the two major areas we analyzed trends in 38 variables separately for each sex. These included the SAT scores and subscores, and 33 SDQ items such as high school GPA, specific course grades, curriculum, numbers of years the students studied various courses, whether or not they were in honors courses, educational plans, ethnic identity, parents' education, and self-estimates of various skills and abilities.

We computed the means and standard deviations of each variable, each year, for males and females separately. In addition, we obtained the intercorrelations of the variables to see which ones were most strongly related to SAT scores. While we could not demonstrate whether there was a causal connection between any particular background variable and subsequent SAT performance, a high correlation with the SAT combined with a similarity in trend would suggest a possible connection. The logic of our analysis was as follows: If a variable is related to SAT scores (causally or otherwise), it will be correlated with SAT scores and it will vary over time in the same way as mean SAT scores vary over time. The trend curves for SAT scores will look similar, therefore, to those for the related variables.

There can, of course, be instances where this generalization does not hold. Patterns of interactions between variables, and correlations of the SAT with unreliable variables, will both attenuate true correlations. Moreover, variables that are truly unrelated may be correlated spuriously because of their mutual relationship with a third variable. The degree to which we accept the interpretation that a variable is causally related to SAT scores will depend to a great extent on the hypotheses we are testing. Generally it will suffice to establish that there appears to be a relationship, and that the relationship may, to some degree, account for the trend or sex difference in SAT scores.

Results of the Analysis of Social Science Majors

With the data from all five years collapsed, we computed correlations of all variables with SAT scores, and we looked at whether the means were significantly different for males and females. Because the sample size was around 5,300, very small differences were statistically significant. Nevertheless, even small differences may be worthy of note because we cannot know at this time how much of a difference is needed to have a significant impact on SAT performance.

Table 3 shows the correlations of each variable with SAT-verbal and SAT-mathematical scores. All of the means that are significantly higher or lower for one sex than the other are indicated by an "M" or "F" in the right column. For example, high school GPA is highly correlated with both SAT-verbal and SAT-mathematical scores. The correlations are .48 and .50, respectively. Furthermore, females have significantly higher GPAs than do males. As we look at the list of correlations, we see that there are moderate to high correlations between SAT scores and many SDQ variables. Most of the ability indicators are correlated. They include GPA, rank in class, grades in English, mathematics, social studies, and physical science, percents in honor courses, whether they plan to apply for exemptions from English courses, and self-estimates of various abilities. Notice that the highest correlate of SAT-mathematical scores (other than other SAT scores) is the single self-estimate of mathematical ability (r = .61) that is higher than the correlation between the SAT-mathematical score and grade in mathematics (r = .44). Similarly the self-estimate of ability in written expression is as highly correlated with the SAT-verbal score (r = .42) as English grade is with the SAT-verbal score (r = .40). These high correlations all suggest that grades and self-estimates of abilities are reliable indicators of academic ability, and though they are not causally related to SAT scores, they are measures of closely related abilities.

When we look at differences in the means of males and females on these same variables, we find some contradictions, however. Males score higher than females on SAT-verbal and SAT-mathematical sections, their average grade in social studies is higher, more are in honors courses, and their self-estimates of mathematics ability, science ability, and ability in written expression are higher than they are for females. Females, on the other hand, have a higher GPA and better grades in English and mathematics, and more females than males plan to apply for an exemption from an English course. We have no basis for explaining these apparent contradictions at this time. We will present some hypotheses in the Discussion.

Table 3. Statistics on Students Planning to Major in Political Science or Sociology, 1978–1982 ($N \cong 5{,}300$)

	Correlations with SAT Scores		*Sex Difference**
	SAT-V	*SAT-M*	
SAT-verbal score	1.00	0.73	M
SAT-verbal reading subscore	0.95	0.70	M
SAT-verbal vocabulary subscore	0.96	0.69	M
Test of Standard Written English (TSWE)	0.80	0.67	
SAT-mathematical score	0.73	1.00	M
High school grade point average	0.48	0.50	F
Percent in nonpublic school	0.05	0.01	
Percent in academic curriculum	0.29	0.27	M
Class size	−0.01	0.01	M
High school rank in class	0.48	0.52	
Number of years studying English	0.18	0.15	
Number of years studying mathematics	0.27	0.42	M
Number of years studying a foreign language	0.38	0.35	F
Number of years studying biological sciences	−0.06	−0.08	
Number of years studying physical sciences	0.23	0.30	M
Number of years studying social studies	0.16	0.12	M
Average grade in English	0.40	0.35	F
Average grade in mathematics	0.32	0.44	F
Average grade in physical science	0.32	0.39	
Average grade in social studies	0.38	0.34	M
Percent in English honors course	0.31	0.28	
Percent in mathematics honors course	0.31	0.39	M
Percent in physical science honors course	0.21	0.24	M
Percent in social studies honors course	0.28	0.27	M
Plan to apply for exemption from English course	0.24	0.19	F
Plan to apply for exemption from mathematics course	0.15	0.30	M
Plan to apply for exemption from physical science course	0.08	0.14	M
Plan to apply for exemption from social studies course	0.14	0.13	M
Hours per week working at a part-time job	−0.04	−0.02	M
Highest level of education planned	0.30	0.31	M
Ethnic identity: Percent white	0.38	0.36	M
Percent for whom English is best language	0.17	0.07	
Highest level of education attained by father	0.35	0.33	M
Highest level of education attained by mother	0.29	0.25	
Self-estimate of mathematical ability	0.36	0.61	M
Self-estimate of ability to organize work	0.17	0.18	
Self-estimate of science ability	0.33	0.40	M
Self-estimate of ability in written expression	0.42	0.31	M

*M = Male average significantly higher than female.
F = Female average significantly higher than male.

From Table 3 we see that all of the SAT scores except for the TSWE have higher means for males than for females. Among the SDQ variables that are correlated with SAT scores, 18 have means that are significantly higher for males than females; only 5 have means that are higher for females than for males. Of these 23 variables, we might expect on a priori grounds that 11 of them would be causal determinants of SAT score performance. These are the following:

1. Percent in an academic curriculum;
2. Number of years studying mathematics;
3. Number of years studying a foreign language;
4. Number of years studying physical sciences;
5. Number of years studying social studies;
6. Percent in a mathematics honors course;
7. Percent in a physical science honors course;
8. Percent in a social studies honors course;
9. Highest level of education planned (an indicator of motivation);
10. Ethnic identity (percent white);
11. Father's education.

Of these 11 variables, the only one that has a higher mean for females than for males is number of years studying a foreign language. While the evidence is not conclusive, it strongly suggests that among social science majors, males come from somewhat better educated families, have higher aspirations, and have a stronger academic education in high school than do females. One would expect these differences in background to account, to some degree, for the sex differences in SAT performance.

The next question is whether any of the 23 variables can account for the trends observed in SAT scores over time. If these background variables are truly indicators of ability, and possibly determinants of SAT score performance, their means should follow the same trends over time as the SAT score means do. Figure 3 shows the shapes of the curves for those ability indicators that correlate most highly with SAT scores (from Table 3). The purpose of the figure is to compare the shapes of the curves visually; the scales are drawn simply as a rough guide, not to be used for accurate reading or to suggest that the scales are comparable from one graph to the next.

The similarity in shape of the eight curves is rather striking. In fact, even males and females have similarly shaped trends. These graphs strongly suggest that the true abilities of this sample declined and then rose to a point very near the 1978 value. If we look at national averages, we see that they followed a very similar pattern during these years, though the SAT-verbal decline for males was ony 5 points and the subsequent rise was only 3 points. Among social science majors, on the other hand, the decline was 25 points and the rise 18 points. Similarly for females and for the SAT-mathematical scores the U-shape of the curve showed a greater decline and greater rise for social science majors than it did for the nation as a whole. Our social science sample also differed in that their SAT-verbal scores averaged some 55 points higher than the national average for males and about 40 points higher for females. SAT-mathematical scores were also somewhat higher than the national average, particularly for males.

With all of these ability indicators suggesting a true decline in verbal and mathematical ability, followed by a true rise, we can look to see what might have caused this trend. A number of background variables from the SDQ were found to be correlated with SAT scores. If these variables also show a U-shaped trend, we will have very strong evidence that they are associated with score changes.

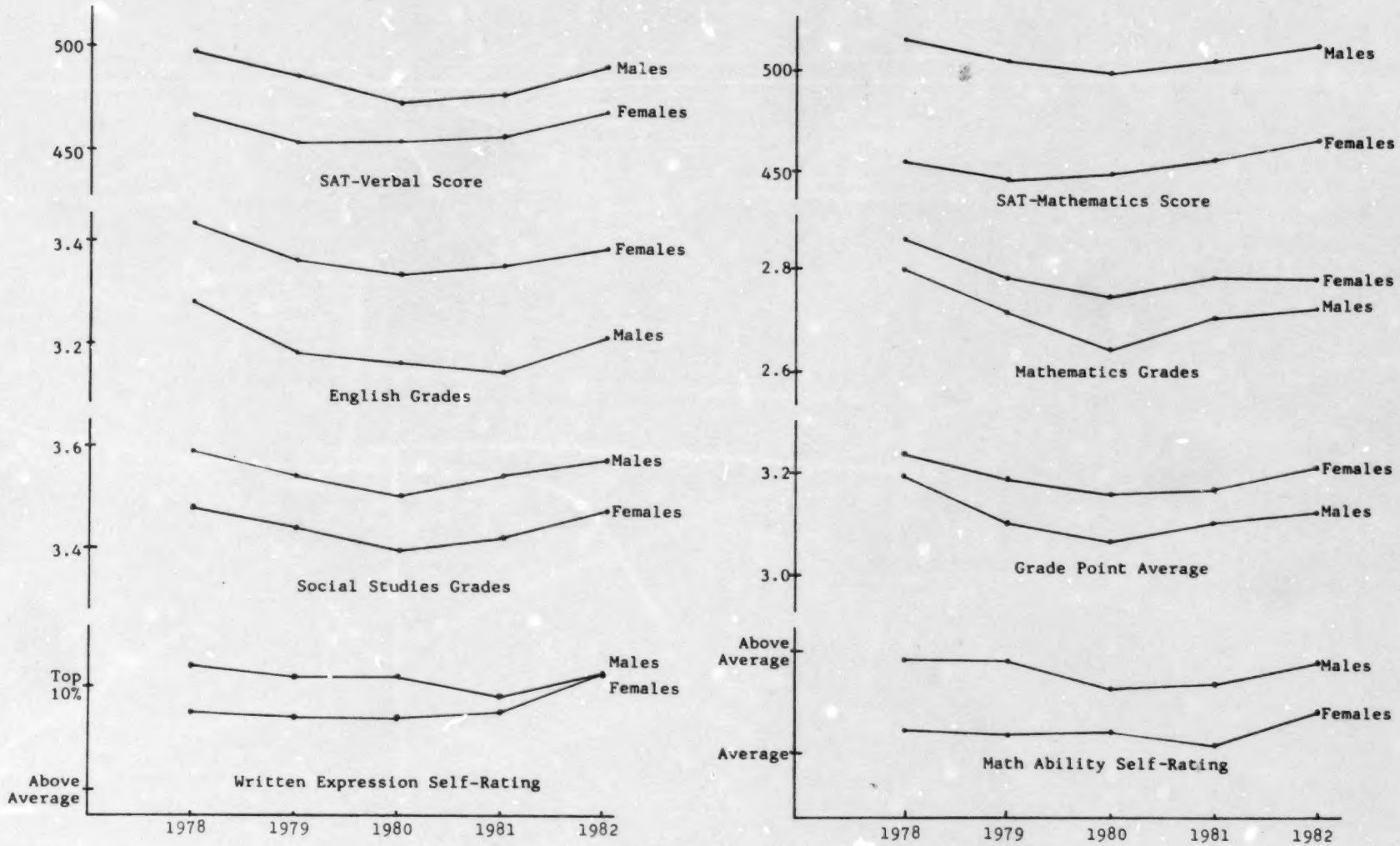
Figure 4 shows trends in six of the background variable highly correlated with SAT scores. These are variables that we might expect to have a causal influence on SAT performance, particularly since they are correlated with it. The curves are not especially U-shaped, however, so it is difficult to conclude that any of these variables are strong determinants of SAT scores. Years of mathematics dips in 1980 for males, in this respect, is similar to the SAT-mathematical score trend. For males, years of mathematics corresponds rather closely to the self-estimate of mathematical ability. For the most part, we cannot conclude that the trends in these background variables can account for the SAT score trends. Quite possibly this may be due to the fact that the mean differences in SAT scores from year to year are just too small to be accounted for by the correspondingly small differences in the means of the background variables. The reliabilities, especially of the SDQ variables, are simply not high enough to account for such small changes in SAT scores. The differences between the sexes, on the other hand, are generally greater than the year-to-year differences within the same sex. Thus the SDQ items may be reliable enough to explain the sex differences. We pose this explanation only as a possibility. To find a more definitive explanation of the score trends we would need to measure these background variables more precisely and do more complex analyses.

Results of the Analysis of Mathematics Majors

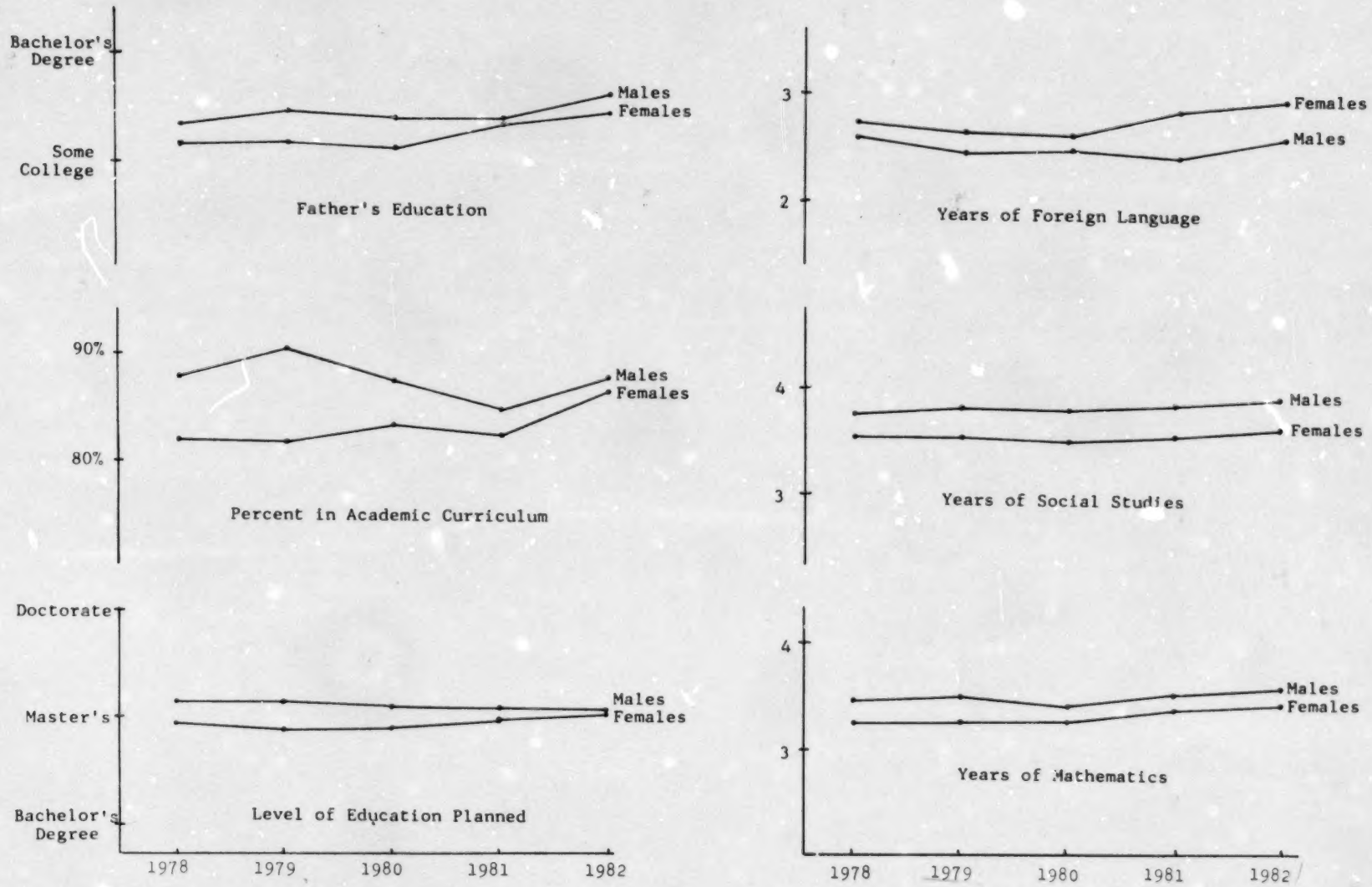
Table 4 shows the correlations of SAT scores with background variables for mathematics and statistics majors. Here we found no sex differences in SAT-verbal scores, and we found females to score higher than males on the TSWE. For the SAT-mathematical scores, however, we still found males to score higher, on the average, than females. Overall, the SAT-mathematical mean was 37 points higher for males than for females. If we compare our sample of mathematics majors with the nation as a whole, we see a great difference in SAT averages. Male mathematics majors score a full 100 points, nearly one standard deviation, higher than males do in general. Female mathematics majors score 114 points higher than the national average for females. While this finding is not surprising, it does confirm our initial assumption that a subgroup of the SAT population having homogeneous interests would be different from the rest of the population.

SAT-verbal scores are also higher for mathematics majors than for the examinee population as a whole, though the difference is only 20 to 40 points. From Table 4 we see males obtained higher means on 9 SDQ variables, and females obtained higher means on 13 variables. This is quite different from the findings for social science majors. Among the indicators of academic ability, males had higher means than females on the following:

1. SAT-mathematical score;
2. Planning to apply for an exemption from a mathematics course;
3. Planning to apply for an exemption from a physical science course;
4. Planning to apply for an exemption from a social studies course;
5. Self-estimate of mathematical ability;
6. Self-estimate of science ability.



18 **Figure 3. Trends in eight ability variables for social science majors 1978–1982.** 19



13 **Figure 4. Trends in six background variables for social science majors 1978–1982.**

Table 4. Statistics on Students Planning to Major in Mathematics or Statistics, 1978–1982 ($N \cong 6{,}600$)

	Correlations with SAT Scores		*Sex Difference**
	SAT-V	*SAT-M*	
SAT-verbal score	1.00	0.73	
SAT-verbal reading subscore	0.95	0.70	
SAT-verbal vocabulary subscore	0.95	0.70	
Test of Standard Written English (TSWE)	0.78	0.68	F
SAT-mathematical score	0.73	1.00	M
High school grade point average	0.49	0.46	F
Percent in nonpublic school	0.06	0.01	
Percent in academic curriculum	0.29	0.27	F
Class size	−0.05	0.00	
High school rank in class	0.50	0.50	F
Number of years studying English	0.14	0.13	F
Number of years studying mathematics	0.25	0.30	
Number of years studying a foreign language	0.33	0.32	F
Number of years studying biological sciences	−0.04	−0.05	
Number of years studying physical sciences	0.29	0.34	M
Number of years studying social studies	0.13	0.09	M
Average grade in English	0.41	0.33	F
Average grade in mathematics	0.29	0.33	F
Average grade in physical science	0.42	0.43	F
Average grade in social studies	0.40	0.33	F
Percent in English honors course	0.29	0.23	F
Percent in mathematics honors course	0.34	0.38	
Percent in physical science honors course	0.25	0.25	
Percent in social studies honors course	0.23	0.19	
Plan to apply for exemption from English course	0.23	0.14	F
Plan to apply for exemption from mathematics course	0.12	0.21	M
Plan to apply for exemption from physical science course	0.16	0.20	M
Plan to apply for exemption from social studies course	0.16	0.10	M
Hours per week working at a part-time job	−0.09	−0.06	M
Highest level of education planned	0.33	0.32	M
Ethnic identity: Percent white	0.32	0.31	
Percent for whom English is best language	0.23	0.11	
Highest level of education attained by father	0.34	0.31	
Highest level of education attained by mother	0.32	0.28	F
Self-estimate of English ability	0.39	0.51	M
Self-estimate of ability to organize work	0.22	0.20	F
Self-estimate of science ability	0.46	0.50	M
Self-estimate of ability in written expression	0.41	0.30	

*M = Male average significantly higher than female.
F = Female average significantly higher than male.

Female examinees, on the other hand, had higher means than males on the following:

1. TSWE;
2. High school GPA;
3. Percent in an academic curriculum;
4. High school rank in class;
5. Average grade in English;
6. Average grade in mathematics;
7. Average grade in physical science;
8. Average grade in social studies;
9. Planning to apply for an exemption from an English course;
10. Self-estimate of ability to organize work.

A comparison of these two lists of ability indicators suggests that female mathematics majors are generally better students than males, as indicated by their grades and possibly by better verbal expression. Their preparation in mathematics, which will presumably be their major, is possibly not as good as the preparation males have had.

Looking at the variables that we would expect to affect SAT performance, this expectation is partially confirmed.

Males have higher means than females on the following:

1. Number of years studying physical science;
2. Number of years studying social studies;
3. Educational plans.

Females have higher means than males on these SDQ items:

1. Number of years studying English;
2. Number of years studying a foreign language;
3. Percent in an English honors course;
4. Mother's education.

The fact that mother's education is very slightly, but significantly, higher for females than for males may account to some degree for the higher proportion of females in an academic curriculum and for the relatively higher academic performance of female mathematics majors than female social studies majors. As we recall, female social studies majors had less well educated fathers than did the males and fewer were in an academic curriculum. This may suggest that females planning to enter mathematics come from a relatively higher socioeconomic status (SES) than do females planning to enter social sciences and that they are somewhat higher in SES than are the males in their field. This higher SES may express itself more in verbal performance than in mathematical performance for females. While we have not attempted to demonstrate a causal connection, these data provide the basis for a working hypothesis for future study.

Figure 5 shows trends in SAT scores as well as other indicators of academic ability for mathematics majors. For the SAT-verbal sections, we see that females actually scored higher on the average than males in 1981 and 1982. Overall, for both sexes there was a decline in scores between 1978 and 1982, though the pattern was clearly not monotonic and may go either way after 1982. The trends for mathematics majors were quite different than the ones we observed for social science majors, and not all indicators followed the same pattern. The trends for males were rather gradual declines with only the slightest tendency to resemble a flattened "S" on its side. The flattened "S" shape was more visible in the female curve for SAT-verbal and SAT-mathematical scores. The only consistent pattern in the trends was that they all declined very slightly over time. Let us see if the correlated background variables declined in a similar way.

Figure 6 shows that mother's education gradually rose a very small amount but remained higher for females than for males. The number of females in an academic curriculum showed a remarkable resemblance to the SAT-verbal and SAT-mathematical score trends for females. For level of education planned, we see, as we did with social science majors, the curves for males and females beginning to converge. The number of years of mathematics study rose for females so that by 1980 it was even with that for males. There was no significant difference overall in the number of years of mathematics studied by males and females. The years of physical science also seem to be converging for males and females.

Our conclusions regarding mathematics majors is that their test scores showed a slight decline overall, both for males and females, and that SAT-mathematical scores were consistently higher for males. Grades were typically higher for females than for males, and they, too, showed a decline. In terms of background variables, we saw a trend toward greater similarity between males and females in SES, aspirations, and coursework.

Discussion

We found, as we expected, that social studies majors and mathematics majors are very different from each other and in most ways are different from the nation as a whole in test scores and educational backgrounds. From our examinations of test score trends and background variables, we concluded tentatively that sex differences in test scores might be related to differences in SES, academic curriculum, and the number of years of study in some academic subjects. Among social science majors, males were higher than females in SES, percent in academic curriculum, and years studying most academic subjects as well as numbers in honors courses. Correspondingly, the males were higher than the females in SAT-verbal and SAT-mathematical scores.

Mathematics majors showed some of these patterns to be reversed. Females were higher than males in SES, percent in academic curriculum, and in course grades. They were at least equal with males on SAT-verbal scores, and they were higher on the TSWE.

What still remains a mystery is why, when there are many indicators that females perform as well as or better than males in academic work, their SAT scores, especially in mathematics, still tend to be lower. One hypothesis is that grades in English and mathematics have a different meaning for each sex. This could occur in one of two ways. First, they may not be pure measures of verbal and mathematical ability. They may also reflect other kinds of performance such as willingness to participate in class discussions, completion of homework assignments, verbal fluency on exams, etc. Psychometrically we would say that grades and SAT scores are not strictly congeneric measures. If we could measure enough background variables, we might be able to dissect grades into their various components to confirm or disconfirm this hypothesis.

Another hypothesis is that a particular letter grade, say a "B," does not correspond to the same level of achievement for males as it does for females. It may be a measure of the same skill for both sexes, but a higher level of skill is required for males than for females to receive the same grade. Psychometrically this is to say that the scales are different for each sex. What it would imply in particular terms is difficult to say. It would seem to suggest that a teacher, consciously or unconsciously, is saying that a male must work harder for his grades.

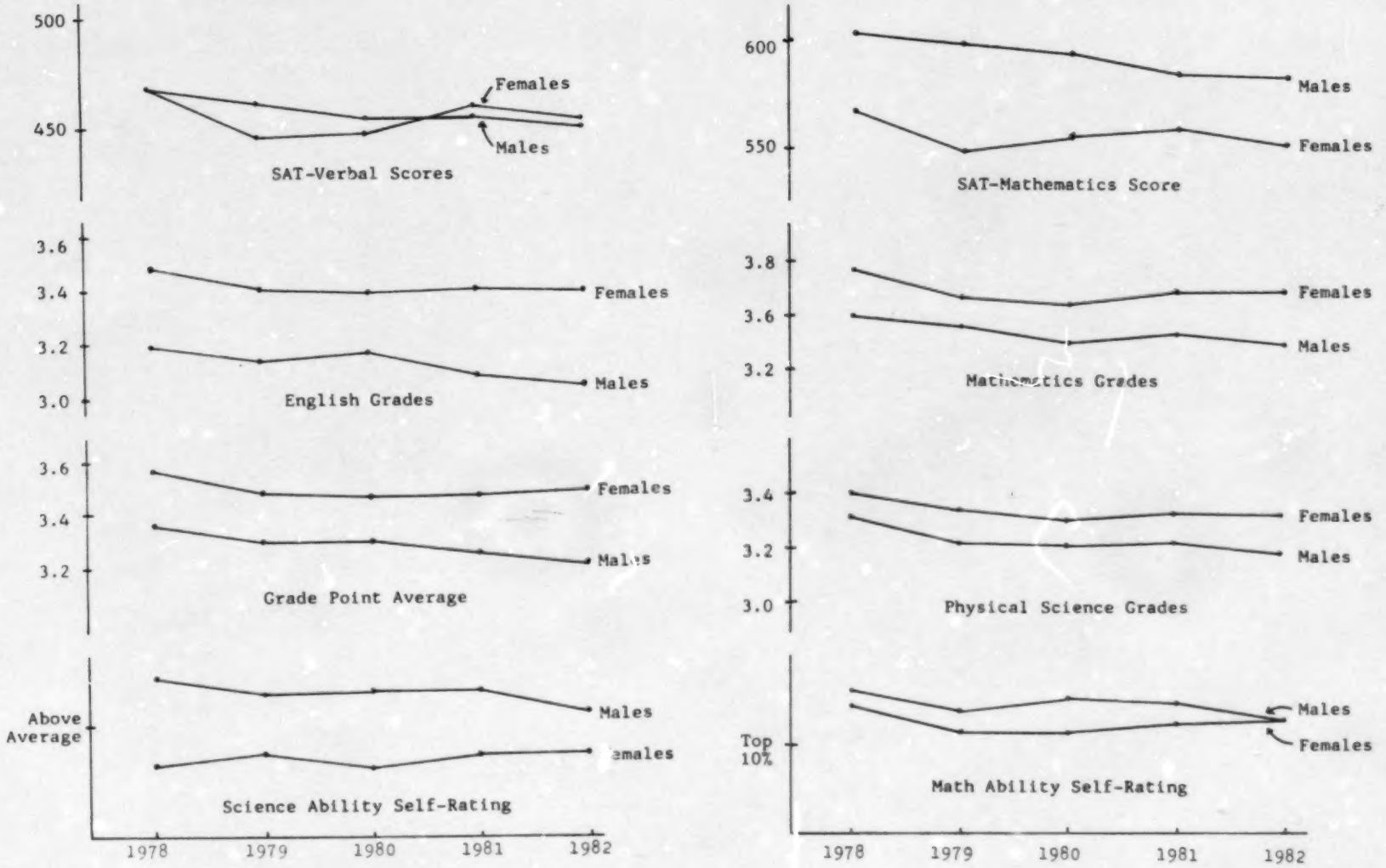


Figure 5. Trends in eight ability variables for mathematics majors 1978–1982.

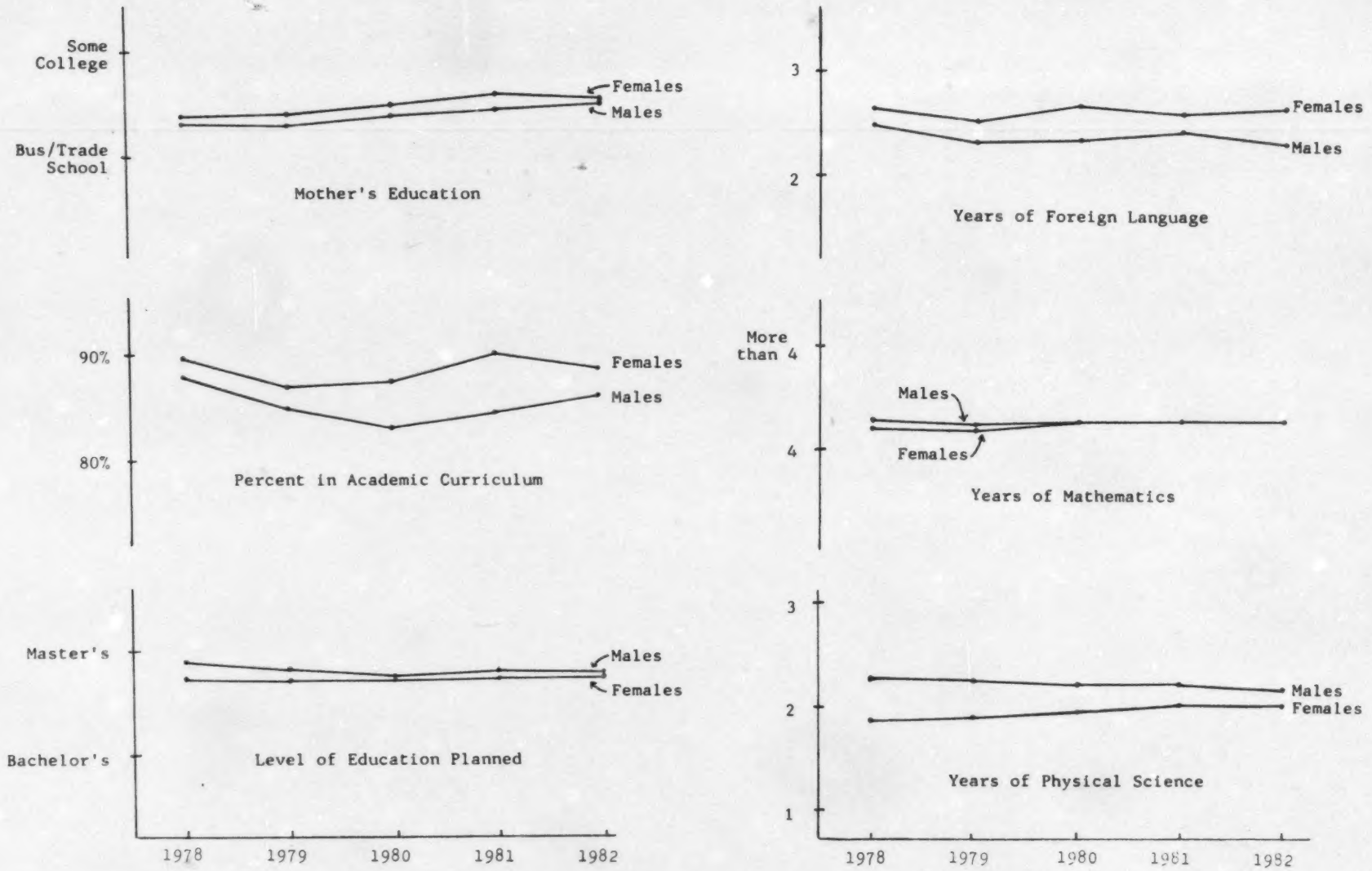


Figure 6. Trends in six background variables for mathematics majors 1978–1982.

Still another hypothesis is that SAT scores reflect more than academic ability. If males are better able to perform under stress, or if they are more "test-wise," or if the content of the items favor male interests and knowledge, their scores would be higher than those of females. It is also possible that the argument for a scale difference applies to the SAT rather than to grades. It is difficult to know exactly what that would mean or why it should occur. Basically it would mean that for the same amount of knowledge and degree of skill, males obtain higher scores on a test that represents a sample of that knowledge. If we could ensure that a test constituted a random sample of knowledge, this would probably not be a problem. But we can never compile a list of all knowledge and sample from it to create a test. Thus, no test is truly a random sample of knowledge. Items are selected in some systematic way which may, inadvertently, favor males. These kinds of hypotheses are not new, and the College Board has been conducting studies to attempt to identify sex biased items.

The present study will not resolve the sex difference problem. What it does suggest, however, is that by studying homogeneous subpopulations of examinees, we can obtain a better understanding of the interrelationships of test scores and other variables without confounding them with changes in students' interests as well. From the two samples we studied, there appears to be a consistency between SES, percent in an academic curriculum, coursework, and grades. It is the SAT scores that appear to be out of line. In one sample, males were generally higher on most of these variables, including test scores. In the other sample, females were higher on most variables *except* test scores. While the case is not strong for saying that there is something inconsistent about SAT scores, they do seem to stand out as the one variable on which females nearly always score lower than males do. It may prove informative to study more homogeneous subpopulations of SAT examinees to see whether any show females to score higher than males. We might then see what background variables can account for the differences.

PREDICTIVE VALIDITY

The previous section examined patterns of test scores and other characteristics of women and men who took the College Board's Scholastic Aptitude Test during the 1970s and early 1980s. This section considers sex differences in academic performance from another perspective, by examining first-year college grades in relation to test scores and high school grades. In this section we ask questions such as "Are college grades equally predictable for men and women?" and "Do admissions criteria such as test scores and high school grades tend to underestimate the college grades that will be earned by women?"

The literature on sex differences in the predictive validity of the SAT has been reviewed by Breland (1978), Linn (1973, 1982), and Wild (1977). In general, previous studies have found that first-year college grades are slightly more predictable for women than for men, and that women tend to earn slightly higher first-year grades than are predicted for them. However, there has not been an investigation of sex differences in predictive validity in recent years. It is possible that, with changing test scores and changing educational practices, indicators of predictive validity also are changing. This investigation was designed to examine the current relationship of precollege measures and college performance as this may vary for the sexes.

Data for the analyses reported in this section were obtained from the file of information from universities that participated in the 1980 College Board Validity Study Service (VSS). Several criteria guided the selection of studies to be included: The data were supplied by a four-year college or university, analyses were done separately for men and women as well as for the sexes combined, and analyses were done for all freshmen or for identifiable specific academic areas of study. In addition, each study had to include SAT-verbal and SAT-mathematical scores and high school grade average or rank in class as predictors, and first-year college grades as the variable to be predicted. The data from 44 colleges and universities met these conditions. Forty-one of these institutions had conducted a validity study for all freshman students (or, in some cases, the bachelor's degree or liberal arts freshmen); 13 of them had conducted a total of 23 studies of freshman students in specific curricular areas. The characteristics of these universities and samples are summarized in Table 5.

As can be seen in Table 5, the colleges and universities that elected to participate in the VSS and that met the criteria of this study are not representative of undergraduate institutions generally in this country. In particular, a high percentage of the studies came from private institutions (more than half) and from universities rather than colleges (two-thirds or more). These characteristics, which reflect the institutions' self-selection to participate in the VSS in 1980 and to study male and female freshmen both jointly and separately, should be kept in mind when reviewing the results of the analyses. Despite these limitations, the studies provide us with a number of independent samples of freshman year academic performance in relation to precollege admission criteria. Therefore, they are useful in addressing questions about the differential predictive validity of admissions variables for men and women, although the results may not be generalizable to all possible conditions in higher education.

Data available for each of the VSS studies summarized below included the number of male, female, and combined sex records included in the analyses, average SAT-verbal and SAT-mathematical scores, average high school grades or rank in class, correlations of the individual and combined predictors with freshman year GPA, and regression weights for the male, female, and combined sex groups. From these data we predicted freshman grades for men and for women using the two-variable (SAT-verbal and SAT-mathematical

Table 5. Characteristics of Universities and Studies in Predictive Validity Analyses

Characteristic	Validity Study of: All Freshmen (N = 41 studies)	Validity Study of: Discipline Freshmen (N = 23 studies)
Institutions:		
Public	44%	35%
Private	56	65
University	63	74
Arts and Science or Liberal Arts College	27	04
Other	10	22
Undergraduate enrollment below 5,000	46	57
Undergraduate enrollment above 15,000	24	30
Tuition below $2,000	44	35
Studies (Sexes Combined):		
Median number of student cases per study	592	176
Average SAT-V score	463	486
Average SAT-M score	499	538
Average H.S. grades	3.06 (14 studies)	2.98 (11 studies)
Average H.S. rank	56.96 (27 studies)	59.72 (12 studies)
Average first-year college grades	2.60	2.71

scores) and three-variable (addition of high school grades or rank) prediction equation for the sexes combined. The average actual GPAs for men and for women were then subtracted from the average predicted male and female GPAs for estimates of over- or underprediction, and the over- or underpredictions were divided by the standard deviations of the earned GPAs for an index of predictive accuracy. Such a measure of effect size has been recommended by Cohen (1977) and used by Hyde (1981) and others in meta-analyses of data in published studies.

The selected VSS studies were summarized in two sets: Studies that included all 1980 freshman students, or all bachelor's degree students, and studies that evaluated predictive validity for freshman students in particular curricular areas such as business, engineering, or biology. The studies of all freshmen are appropriate because, even though the distribution of men and women across major fields is quite different, there is probably some commonality in the courses taken by men and women college freshmen to satisfy general university curricular requirements. However, we would expect even more commonality in the freshman courses of men and women in the same curricular field; therefore, the discipline samples are of particular interest because they minimize even further the likelihood that different courses and grading practices are represented in the male and female freshman grade averages.[1]

1. Wild and Dwyer (1977) and Dwyer (1979) reviewed the problems of using teacher-assigned grades and grade-point averages as criterion scores in the analysis of the predictive validity of test scores. The effects of field differences in grading are illustrated in a study by Goldman et al. (1974); Goldman and Slaughter (1976) noted the nonequivalent nature of grades that are used to compute a composite GPA, resulting in unreliability in the criterion that will limit the potential sizes of correlation coefficients.

In general, the pattern of average test scores and grades in the two sets of studies are similar to the pattern of precollege data presented earlier, with higher mathematics test scores for men than women and higher earned grade averages (both in high school and in college) for women. However, the average SAT-verbal score is higher for women than for men in the all-freshmen groups (463 for women, 451 for men), which reverses the male advantage (of eight points) found among all 1980 college-bound seniors (*National College-Bound Seniors* 1983, p. 4). This shift may reflect differences in individual choices to attend and complete the freshman year in college, or may be a distortion introduced by the self-selected nature of institutional data in the VSS file. In any case, it seems to be characteristic of VSS data, where the average verbal scores of women were higher than the verbal scores for men in each year from 1964 to 1978 (Bejar and Blew 1981, p. 5).

Correlations with First-Year Grades

One way to consider the predictive validity of precollege measures is to examine the correlations between these measures and college performance. Table 6 presents a stem-and-leaf tabulation of multiple correlation coefficients between predictors and first-year grades in 41 studies of 1980 freshmen. When only test scores are used (the two predictor data, only available in 30 of the 41 studies), the median correlation with the first-year grades of women is .46 which is higher than the comparable correlation of .38 for men. But when high school record is added to the multiple relationship, the predictive advantage for women is reduced considerably to median correlations of .52 for women and .51 for men. Almost identical medians have been reported for more than 200 VSS studies conducted between 1976 and 1980 that were classified by sex, except

Table 6. Multiple Correlations Between Predictors and First-Year College Grades in 41 Validity Studies of 1980 Freshmen Students

Correlation	*Males*	*Females*
Two Predictors ($N = 30$)[a]		
.1		
.2	3467	5
.3	2334566788899	01337799
.4	012347899	11234567899
.5	1237	002268
.6		4588
Median	.39	.46
Mean	.39	.45
Three Predictors ($N = 41$)[b]		
.1		
.2	0	
.3	133467	168
.4	123345778899	224456678899
.5	112233346688889	001122236678899
.6	0113788	01356999
.7		124
Median	.51	.52
Mean	.50	.54

Note on reading the table: The "stem" or decimal number in the left column is the first digit of a correlation coefficient; the "leaf" or each number in a "bar" is the second digit of the correlation. Thus, the table gives both a graphic and detailed frequency tabulation of the obtained multiple correlation coefficients in this set of validity studies.
[a]Two predictors = SAT-V and SAT-M scores.
[b]Three predictors = SAT-V and SAT-M, and high school GPA or rank.

that in this larger set of studies the median three-predictor multiple correlation with first-year grades was .57 for women and .52 for men (*ATP Guide* 1983-84, p. 22). Both sets of VSS results are consistent with the median multiple correlations summarized by Breland (1978, p. 38) from more than 50 published validity studies. As reported by Breland, Ford and Campos (1977), and others, the combination of test scores and high school record provides the highest median correlation with freshman grades, and the first-year college grades of women tend to be slightly more predictable (higher median correlation) than the first-year college grades of men.

A recent study to examine the phenomenon of college grade inflation and the validity of the SAT (Bejar and Blew 1981) examined studies in the College Board VSS file over a 15-year period, 1964 through 1978. All results were reported separately for males and females. The average three-predictor multiple correlation coefficients for men ranged from .55 in 1964 to .51 in 1978, while the average multiple *R* for women ranged from .65 in 1964 to .55 in 1978. These data suggest that college freshman grades have become slightly less predictable by test scores and high school record over the last 20 years, and that the grades of women have consistently been slightly more predictable than the grades of men. The mean three-predictor multiple correlations obtained in the current study (see Table 6) of .50 for men and .54 for women are consistent with these trend data.

Under- and Overprediction

A second and perhaps more important way of considering the predictive validity of precollege measures is to compare actual first-year college grades with the grades that would be predicted by the admissions variables. When used in admissions, a regression equation based on one year's data is used to predict grades for another year's applicants, thus assuring the independence of the data that are used to generate the equation and the data that will be used to make predictions. In this study, we are concerned instead with the question of whether or not precollege measures predict the grades of men and women with equal accuracy. Therefore, the three-predictor regression equation for the total group of students in each study was used to predict first-year grades separately for men and for women, and these predicted grades were compared with the grades actually received by each group. The results of this analysis are presented in Table 7.

Data from the 23 validity studies in specific disciplines are presented individually in Table 7, since they represent the most homogeneous samples of male and female students. For example, men and women studying engineering are probably much more alike both in academic backgrounds and in freshman course enrollments than are male and female college freshmen generally. Therefore, the likelihood of similar prediction systems is maximized. Average data across the 23 discipline studies are provided near the end of the table, together with average data for the 41 studies of entering freshman classes.

As can be seen in Table 7, first-year college grades are slightly underpredicted for women and overpredicted for men when a common regression equation is used to predict the grades. As expected, the amount of underprediction is less in the more homogeneous discipline groups than for all freshmen, but grouping students by disciplines does not entirely remove the sex difference. It should be noted, however, that the direction of overprediction is reversed in 5 of the 12 scientifically oriented studies in Table 7, and in 2 of the 6 business studies, but not in any of the 5 humanities-oriented studies. The number of cases is too small to draw any conclusions, but these results suggest that the small variations in predictive validity between the sexes, such as are reported here, may be reflecting relatively random differences in the courses taken by men and women, or in the assignment of course grades. It is interesting to note that, almost always, the actual first-year grade average is higher for women than men when the women's grades are underpredicted, and the actual grades usually are higher for men than for women when the women's grades are overpredicted. Linn and Werts (1971) have shown that the

Table 7. Accuracy of First-Year College Grade Predictions for Groups of Men and Women in the 1980 College Board Validity Study Service

	Men					Women				
Group Predicted	*N*	*Predicted[a] Grades*	*Actual Grades*	*Difference[b]*	*Difference[c] Index*	*N*	*Predicted[a] Grades*	*Actual Grades*	*Difference[b]*	*Difference[c] Index*
Engineering	105	1.98	1.96	.02	.02	13	2.06	2.18	−.12	−.15
	73	2.92	2.91	.01	.01	74	2.78	2.79	−.01	−.03
	210	2.65	2.61	.04	.04	40	2.61	2.80	−.19	−.28
	494	2.64	2.65	−.01	−.02	125	2.48	2.40	.08	.13
	263	2.56	2.56	.00	.01	71	2.71	2.73	−.02	−.03
Pharmacy	44	2.48	2.66	−.18	−.23	67	2.57	2.45	.12	.14
Health Science	46	2.89	2.93	−.04	−.07	91	2.79	2.77	.02	.03
Biology	149	2.83	2.86	−.03	−.04	182	2.63	2.62	.01	.02
	59	2.81	2.83	−.02	−.03	48	2.76	2.74	.02	.04
Science Majors	223	2.69	2.69	.00	.00	73	2.61	2.61	.00	.00
Science and Math	53	2.93	2.83	.10	.15	112	2.93	2.98	−.05	−.09
Agriculture and Home Economics	79	2.52	2.50	.02	.02	140	2.59	2.60	−.01	−.01
Business	55	2.90	2.83	.07	.12	53	2.99	3.06	−.07	−.14
	95	2.76	2.80	−.04	−.07	99	2.82	2.79	.03	.05
	114	3.11	3.10	.01	.03	85	3.05	3.07	−.02	−.04
	69	2.56	2.59	−.03	−.05	61	2.66	2.63	.03	.04
	210	2.62	2.62	.00	.00	227	2.60	2.60	.00	.00
	311	2.45	2.34	.11	.12	254	2.56	2.70	−.14	−.16
Liberal Arts	406	2.53	2.50	.03	.04	510	2.63	2.65	−.02	−.03
Humanities	62	2.88	2.80	.08	.12	114	2.86	2.90	−.04	−.09
Speech and Drama	22	2.81	2.65	.16	.30	49	2.85	2.93	−.08	−.13
Communications	92	2.74	2.68	.06	.10	68	2.76	2.84	−.08	−.12
Fine Arts	45	3.04	2.97	.07	.12	62	3.05	3.10	−.05	−.11
Average 23 Depts.	143	2.71	2.69	.02	.03	114	2.71	2.74	−.03	−.04
Average 41 Freshman Classes	817	2.57	2.52	.04	.06	845	2.64	2.68	−.05	−.06

[a] Based on three variable prediction equation for total group (SAT-V, SAT-M, and high school grade average or rank).
[b] Predicted minus actual grades, or amount of overprediction.
[c] Overprediction divided by standard deviation of GPA.

lower scoring group will be overpredicted when the measurement reliability of the predictors is less than perfect; omitting a predictor variable which is related to the criterion and on which there are group differences may also lead to overprediction for one of the groups (Linn 1982).

The last column in each set of data in Table 7 reports a "difference index" created by dividing the predicted versus actual grade difference by the standard deviation of the actual college GPA (Cohen 1969). This measure indicates the percent of a GPA standard deviation that is represented, on average, by the under- or overprediction. For women in the discipline studies, seven of 23 sets of actual grades were underpredicted by 10 percent of a GPA standard deviation or more, while two of 23 average grades were overpredicted by a similar amount; the highest underprediction was 28 percent of a GPA standard deviation in an engineering program made up 5 to 1 of men. The average underprediction for women in the discipline studies was 4 percent of a GPA standard deviation, while the average underprediction for women in the 41 freshman class studies was 6 percent of a GPA standard deviation. Cohen (1969) considers a difference ratio of .20 to be small, a value of .50 to be medium, and a value of .80 to be large. Based on these suggested levels, the difference ratios observed in this study must be considered very small.

These results come as no surprise, since the underprediction of women's first-year college grades has been reported consistently in the research literature (see, for example (American College Testing Program 1973; Breland 1978; Novick 1982; R L. Linn 1973 and 1982)). In each of these analyses, the sex differences in prediction were greatest when individual predictors were used, and least

when a combination of test scores and high school grades were included in the regression equation. This same pattern was observed in the current analysis, where women's grades were underpredicted by 10 to 15 percent of a GPA standard deviation when only test scores were used in the prediction formula (data not shown), compared to 4 to 6 percent underprediction (reported in Table 7) when both test scores and high school record were used. As suggested by Wild (1977), the underprediction does appear to be minimized further when combinations of predictors are used for the prediction of performance in specific major field areas, but the sex differences are not entirely eliminated. This seems to be true even when single course grades are used as the criterion, as found in nine VSS studies analyzed for grades in English or mathematics courses as evidence of underprediction for women (Donlon, in press).

In summary the underprediction of women's college grades seems consistent and pervasive, but generally the bias is at a low level of magnitude when multiple measures are used in the prediction equation and male and female groups are selected to be reasonably similar in course enrollments. When the male and female groups are fairly different in course enrollments, as is probably true for analyses of total freshman classes, it is reasonable to speculate that the higher average first-year grades of women may reflect higher levels of enrollment in courses that tend to be graded somewhat more leniently than are the quantitatively oriented courses that tend to be favored by men (Goldman et al. 1974). This explanation is less persuasive for students in the same major field or even the same course, although bias favoring women in the assignment of grades cannot be ruled out entirely.

Linn (1982) suggests that more careful selection of variables can reduce gender differentiated prediction. One of these variables may be writing ability, since test scores have also been found to underpredict women's scores on the Test of Standard Written English writing samples (Breland and Griswold 1981) and writing skill is an important aspect of grading in many college courses. Another possible solution would be to use separate prediction equations for men and women. However, a recent cross-validation study by the American College Testing Program on the accuracy of freshman grade predictions (Sawyer and Maxey 1982) concluded that "there is typically little or no benefit in calculating separate-sex multiple regression equations. Some other prediction method, however, such as a combined-sex equation with adjusted intercept, might offer improved prediction" (p. 6).

Is the underprediction of first-year grades for women sufficient to be of general concern? One must respond affirmatively in view of the pervasiveness of the evidence of underprediction especially when only one or two variables are used to make the prediction. However, when multiple measures are used for groups with reasonably similar backgrounds and interests, the gender differences are very small—generally less than 10 percent of a GPA standard deviation. In view of the small size of the variations, it seems likely that the phenomenon of differential prediction for men and women is of more concern to statisticians than to university students or administrators.

CONTENTS AND STRUCTURE OF THE SAT

Although the panel of experts appointed to investigate the SAT score decline in the 1960s and early 1970s (*On Further Examination* 1977) found no evidence that changes in the contents or structure of the tests were related to observed declines in both verbal and mathematical skills for women and men, there has been a continuous stream of studies over the past 15 years to evaluate possible sex bias in the tests themselves. These investigations have taken several forms; most frequently, they have considered: (1) sexist language in test items or differences in the contents of stimulus materials such as reading passages that might give an advantage to members of one sex; or (2) gender-differentiated performance on items classified by type (for example, analogies or geometry) or level of difficulty; or (3) the equating of scores on one form of a test with scores in previous years as well as scores on new forms; or (4) psychometric issues such as speededness, item omissions, distractor analysis, and scoring methods. While a complete review of this literature is beyond the scope of the present paper, this section summarizes the current state of knowledge on some of these issues.

Verbal Performance

A great deal of attention in recent years has been given to the persistent sex difference in performance on the SAT-mathematics measure on which men consistently average scores 40 to 50 scaled score points higher than women. However, very little attention has been given to the shifting scores on the verbal measure, where women's scores were slightly higher than men's until 1972 but now average 10 points lower than men.

Though it has not been possible to identify specific changes in the SAT-verbal sections that would account for the relative shift in mean scores from higher for females to higher for males, there were a number of extensive studies of the test in the late 1960s and early 1970s that were prompted by the general decline in scores (see (Donlon, in press) for a review of these studies). Also, there have been changes in the contents of reading passages and items to comply with requirements of the Educational Testing Service (ETS) test sensitivity review process (Hunter and Slaughter 1980). It seems possible that a number of small adjustments may have had unexpected and undocumented cumulative effects relative to the sex-related verbal score shift. The likelihood of such an occurrence has been suggested by Tittle, McCarthy, and Steckler (1974) and by Dwyer (1979), and may result in part from conscious efforts

to "balance" the test to be sex-neutral. In the process, it appears that the relative slight advantage has been shifted from women to men.[2]

Mathematical Performance

Sex differences in performance on the mathematical sections of the SAT have been observed since the earliest introduction of mathematical material in the 1930s (Donlon, in press). However, only with the advent of SDQ data has it been possible to compare scores of males and females who report the same number of mathematics courses in high school. Table 8 presents mean test scores by self-reported years of mathematics study among college-bound seniors in 1978–79. The progressively higher average test scores for both men and women indicate that continued study of mathematics results in greater skill in this area. However, the average gender score differences appear to be least for students who report 2 or 3 years of mathematics study; men and women who report more than 4 years of mathematics study in high school appear to have one of the largest score differentials, with a difference of 50 points between the average men's score of 585 and the average women's score of 535. Unfortunately, there is no way of knowing whether the male and female test takers are reporting courses of similar content and difficulty in high school. Given the difference in scores, it seems very possible that women are reporting a greater number of less advanced mathematics courses than are the men, thus contributing to a false sense of fair comparison when, in fact, the mathematical instruction and experience of the two groups continues to be fairly different. If this hypothesis is correct, the difference in training is even greater than the 65 percent of males versus 45 percent of females with four years or more of high school mathematics, as reported in Table 8.

Unfortunately, data on the standard deviations of male and female SAT-mathematical scores are not available for the groups identified by years of prior mathematics study. In general, the SAT-mathematical standard deviations of college-bound senior women are slightly smaller than the score variations of college-bound senior men (105 to 109 for women versus 114 to 120 for men over the past 15 years). The slight increase in score variation for both sexes may be related to increasing variation in the prior mathematical training of test takers (Donlon, in press, Table 9.3).

Questions about the usefulness of the differential socialization and experience explanations for the male–female gap in levels of quantitative performance have been raised by Benbow and Stanley (1982), based on a five-year follow-up of over 2,000 seventh and eighth graders who scored sufficiently high on the SAT to be included in a Study of Mathematically Precocious Youth (SMPY). In this study, the sex difference persisted over several years and was related to subsequent sex differences in mathematics achievement, but the difference did not reflect differential mathematics course taking in junior and senior high school or in expressed attitudes toward mathematics. However, because these talented males reported taking their mathematics in significantly earlier school grades than females, significantly more males than females were able to take calculus and other advanced mathematics courses in high school.

A more direct test of the hypothesis that the sex difference in quantitative SAT performance may be due to differences in the pattern of quantitative coursework taken by males and females in high school is reported by Pallas and Alexander (1983). Based on a more broadly representative sample, their study suggests that the male–female gap in SAT-mathematical performance shrinks considerably when differences in quantitative high school coursework are taken into consideration. They report that when differences in quantitative coursework were controlled in their analyses, the male–female gap in SAT-mathematical performance decreased from 35 points to 14 points. Females tended to outperform males in the quantitative courses they took together.

Though analytically persuasive, data for the Pallas and Alexander analyses came from the ETS growth study with data collected from 1961 to 1968; it seems possible that more recent advancements in the teaching of higher level mathematics at the high school level, coupled with greater awareness of the need to encourage young women to take more mathematics courses, could account for the discrepancy in their results relative to the results of Benbow and Stanley.

A recent study using data drawn from the 1977–78 National Assessment of Education Progress in mathematics (Welch, Anderson, and Harris 1982) found that 25 percent of the variance in mathematics achievement was accounted for by background variables such as characteristics of the community and educational level of parents, while the number of semesters of mathematics studied explained an additional 34 percent of the variance. The results were replicated on three different national samples of 17-year olds. These results, based on a broader population and using a test other than the SAT, support the general conclusion that sex-related differences in performance on tests of mathematical skill reflect formal coursework in mathematics, but a score difference favoring males remains after the effects of differential preparation have been removed.

In a study of a related field, Wheeler and Harris (1981) examined sex differences in performance on the ATP physics test. Although numerous studies have documented the lower levels of participation by women in science courses and the higher test scores of men in the sciences, there have been few studies of the sources of these differences. In a detailed analysis of item types as well as test takers' backgrounds and interests, Wheeler and Harris determined that test score

2. The classic example of balancing test content for sex is the Stanford-Binet intelligence scale, when the 1937 revision introduced more items favoring males in order to reduce a small but stable superiority favoring females in the 1916 version (Dwyer 1979, p. 342).

Table 8. Comparative Averages for Males and Females on the Mathematical Sections of the SAT by Years of Preparation (Based on SDQ Results, 1978–79)

Years of Study	Males		Females		Difference in Average Score
	N	SAT-M Average	N	SAT-M Average	
>4	50,735	585	30,491	535	50
4	225,344	520	180,252	482	38
3	101,607	449	167,116	427	22
2	38,224	395	77,318	376	19
1	6,912	376	11,353	346	30
0	1,465	393	2,064	354	37
Total	424,287	496	474,594	445	51

From (Donlon, in press).

differences were relatively consistent across item types, number of semesters of mathematics, and high and low achievement bands. However, there were only three items of 75 on the test for which the proportion of women passing the item equaled or exceeded the proportion of men passing the item. The only unusual finding was that female candidates omitted almost 19 percent of the items they reached, while male candidates omitted only 11 percent of the items; the authors speculated that further investigation of the effects of omitting items might be profitable in explaining gender-related score differences in mathematics and the sciences.

Donlon (1973) has shown that the mathematics test scores of women are influenced by the contents of the test, to the extent that the SAT-mathematical sex differences could be reduced by about 20 scaled score points if the content of the test were focused primarily on algebra, or inflated to about 60 points if it were focused on geometry. The data sufficiency item type also has been shown to favor females; however, this item type has been dropped from the SAT.

To summarize, the long-held supposition that male–female differences in scores on the mathematical sections of the SAT reflected different levels of mathematical knowledge gained through previous coursework and other life experience appears to be only partially true. Most studies do find that the continued study of mathematics results in higher test scores, and that males tend to pursue mathematics longer and in greater depth than do females, but controlling for years of study does not appear to explain completely the differences. Benbow and Stanley (1982) suggest that the differences may be inherent sex differences; others, such as Hyde (1981), Linn et al. (1981), and Wittig and Petersen (1979), emphasize the interaction of natural capacities with environmental opportunities. Whatever the causes of sex differences in the demonstrated mathematical skills of college-bound seniors, the literature that was reviewed in the section "Literature on Cognitive Functioning" on sex differences in cognitive functioning and in the section "Predictive Validity" suggest that the differences account for very little of the variance in academic performance. Clearly, young women close off many occupational opportunities when they fail to obtain solid grounding in mathematics and science prior to college; the importance of continuing to encourage young women to enroll in mathematics and science courses cannot be overemphasized. But we should not expect that these steps alone will be sufficient to equalize the average male and female performance on the SAT-mathematical measure.

Test Construction and Administration

Dwyer (1979) has reviewed the implications of the practices used in constructing tests, such as the SAT, that will be used to identify or interpret sex-related differences in cognitive functioning. She notes that test specifications generally define the contents of the test in some detail, both in terms of the types of items to be included and the desired technical characteristics of the final form of the test. In recent years, there has also been a review for test sensitivity concerning sexist language and contents (Hunter and Slaughter 1980). Decisions about test contents generally are matters of judgment rather than empirical efforts to identify bias; the possible implications for gender-differentiated performance rarely are considered.

An important aspect of test construction is a determination of the number of items that will be administered in a given period of time. This leads to a concern for speededness, often defined in terms such as "most students should respond to at least 75 percent of the questions," and "at least 80 percent of the students should reach the last question." Kappy and Donlon (1977) found little evidence to support an overall sex-related difference in the rate of work on moderately speeded cognitive tests. However, they recommended further research in which ability groups and subject matter areas are analyzed separately.

Differential test performance depending upon item content has been mentioned earlier in this section. The *context* of a test item, as well as its contents, has been found to make a difference in the performance of females and males (Dwyer 1979). That is, items or reading passages concerned with literary figures or human relationships tend to be somewhat easier for females; items concerned with technology or the world of public affairs tend to be somewhat easier for males (Strassberg-Rosenberg and Donlon 1975). As Dwyer points out, it is more common for test developers to attempt to assess the sex-relatedness of context rather than content, "but the process may still be best described as *ad hoc*" (1979, p. 346).

Standard ETS procedures for examining the possible bias of individual test items have been to look for item-group interaction in the difficulty of an item (Angoff and Ford 1973) or to compare item characteristic curves for the groups under investigation (Lord 1977). Many of these studies have been reviewed recently by Lockheed (1982). Generally, few individual items on the SAT have been found to be associated with differences in the performance of women and men; when items have been identified, they have not been of any particular type or content. However, there are continuing efforts to find new methods of identifying sex bias in the tests. For example, Donlon (1982) has been looking at differential patterns in the choices of distractors (errors) on different types of items, and Dorans and Kulick (1983) have developed a "standardization" approach to control for differences in subpopulation ability before making performance comparisons on test items.

The last aspect of test construction that Dwyer (1979) examined is concerned with the comparability of alternate and successive forms of a test, or comparability across age levels. Clearly, such psychometric considerations are closely related to item content, and yet test equating typically takes place without regard to the content of specific items or changes in the content specifications for test construction. Thus, changes in content that have implications for sex-related performance may occur from form to form, but the differences will be masked when the forms are equated to produce comparable scores for the gender-combined pools of test takers.

As suggested earlier in discussing the possibility that small changes in the verbal test may have contributed to a larger average SAT-verbal score decline for women than for men over the past 15 years, the exact impact of variations in the test contents, structure, and administration are very difficult to determine. Almost certainly, some of the changes that have been made in the tests have altered the potential performance possibilities for individual students. At the same time, it is apparent that the likely impact of the changes is very small relative to the substantial similarity of performance by women and men. Efforts should be made to reduce the irrelevant impact of testing procedures on academic assessment, but large changes in scores should not be expected as a result of these activities.

SUMMARY AND CONCLUSIONS

This report has reviewed a fairly large body of literature that describes the personal characteristics and test scores of college-bound senior men and women who have taken the Scholastic Aptitude Test (SAT), examined theories and evidence concerning the nature and causes of sex differences in cognitive functioning, and considered the evidence for sex-related test bias either in the test itself or in the use of test scores to predict first-year college academic performance. In addition, three empirical investigations are reported. In the first, the characteristics of the SAT examinees in 1973 and 1982 are compared to look for changes over this 10-year period in family backgrounds, career interests, and high school preparation that might help to explain the greater test score decline for women than for men on both the verbal and mathematical tests.

The second study selected two more homogeneous groups of college-bound seniors, those who said they were interested in majoring in political science or sociology and prospective mathematics or statistics majors, and compared the characteristics of male and female test takers for the years 1978 to 1982 as a way of looking for systematic similarities and differences between the sexes.

The third study examined the correlation between test scores and college performance separately for men and women, and compared actual first-year college grades with the grades predicted by test scores and high school records.

The results of the empirical studies conducted for this report are consistent with the results of similar studies in the literature. The test scores of women have declined more than the scores of men over the past 15 years, and the characteristics of the test-taking groups have changed, but it is not clear that the demographic changes account for the score declines. Using both test scores and high school records to predict first-year college grades continues to work reasonably well for both men and women. Studies consistently demonstrate slightly higher correlations between the predictors and college grades for women; however, the first-year grade averages of women tend to be underpredicted slightly, perhaps because women frequently average higher college grades than predicted.

The literature on sex differences in cognitive functioning consistently reports a slight female advantage in the verbal area (increasingly not true over the last 15 years for SAT-verbal scores) and a somewhat larger but still small male advantage in the mathematical area (consistently true for SAT-mathematical scores). However, increasingly this literature is emphasizing the small amount of the variance in cognitive ability that is accounted for by gender—generally, no more than 5 percent. Therefore, although the causes of the observed differences continue to be investigated, the emphasis is upon the much greater degree of similarity between the sexes in cognitive ability.

The question of test bias, either in the test itself or in the use of test results, continues to be of interest. As has

been reported in other studies, this investigation found a slight underprediction of first-year college grades for women, which could be interpreted to mean an admissions handicap for women. However, the underprediction was very small when both test scores and high school records were used to predict grades for reasonably homogeneous groups—generally less than 10 percent of a GPA standard deviation—and the women almost always averaged higher earned grades than did the men. Therefore, it seems likely that the basis of the differential prediction rests more in the criterion than in the predictors, and that the problem is not a very important one for college admissions.

The possible impact of changes in the tests on test taker performance is harder to document, but it seems likely that some of the changes in item content and format, perhaps even some of the changes made in an effort to assure "sex-neutral" tests, have had some effect on scores, and that these effects have been felt more by women than by men. In any event, it appears to be a possibility that deserves further consideration.

This examination of literature and data has not answered the questions that prompted the study, especially "Why have women's SAT test scores declined more than men's test scores over the past 15 years?" and "Why do women consistently score lower than men on both the verbal and mathematics tests?" It is hoped that it has, however, provided a better understanding of the complexity of these questions. It has also demonstrated the clearer results that can be obtained when samples of men and women are chosen to be as homogeneous as possible, such as in the same major field of study. This was true for the investigation of predictive validity as well as for score decline, where differential prediction of grades was minimized in major field groups. More careful attention to sampling in future studies could contribute considerably to the usefulness of the results.

REFERENCES

American College Testing Program. 1973. Bias in the Uses of Educational Assessments. In *Assessing Students on the Way to College* (Volume I). Iowa City, Ia.: Author.

Angoff, W. H. and Ford, S. F. 1973. Item-Race Interaction on a Test of Scholastic Aptitude. *Journal of Educational Measurement* 10: 95–106.

ATP Guide for High Schools and Colleges (1983-84). New York: College Entrance Examination Board.

Benbow, C. P. and Stanley, J. C. 1982. Consequences in High School and College of Sex Differences in Mathematical Reasoning Ability: A Longitudinal Perspective. *American Educational Research Journal* 19: 598–622.

Bejar, I. I. and Blew, E. O. 1981. Grade Inflation and the Validity of the Scholastic Aptitude Test. *American Educational Research Journal* 18: 143–156.

Breland, H. M. 1978. *Population Validity and College Entrance Measures* (College Board Research and Development Report 78-79, No. 2). New York: College Entrance Examination Board.

Breland, H. M. and Griswold, P. A. 1981. *Group Comparisons of Basic Skills Measures* (College Board Report No. 81-6). New York: College Entrance Examination Board.

Casserly, P. L. 1982. *Older Students and the SAT* (College Board Report No. 82-8). New York: College Entrance Examination Board.

Cohen, J. 1977. *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.

Deaux, K. 1984. From Individual Differences to Social Categories: Analysis of a Decade's Research on Gender. *American Psychologist* 39: 105–116.

Donlon, T. F. 1982. Comparison of Male and Female Response Patterns on Error Choices in Aptitude Test Items as Evidence of Test Fairness. Presented at the New England Educational Research Organization, April.

———. 1973. *Content Factors in Sex Differences on Test Questions* (Research Memorandum 73-28). Princeton, N.J.: Educational Testing Service.

Donlon, T. F., ed. In Press. *The Technical Handbook for the College Board Scholastic Aptitude Test and Achievement Tests*. New York: College Entrance Examination Board.

Dorans, N. J. and Kulick, E. 1983. Assessing Unexpected Differential Item Performance of Female Candidates on SAT and TSWE Forms Administered in December 1977: An Application of the Standardization Approach (ETS Research Report No. 83-9). Princeton, N.J.: Educational Testing Service.

Dwyer, C. A. 1979. The Role of Tests and Their Construction in Producing Apparent Sex-Related Differences. In *Sex-Related Differences in Cognitive Functiong*, Wittig and Petersen, eds. New York: Academic Press.

Ford, S. F. and Campos, S. 1977. *Summary of Validity Data from the Admission Testing Program Validity Study Service*. New York: College Entrance Examination Board.

Goldman, R. D., Schmidt, D. E., Hewitt, B. M., and Fisher, F. 1974. Grading Practices in Different Major Fields. *American Educational Research Journal* 11: 343–357.

Goldman, R. D. and Slaughter, R. E. 1976. Why College Grade Point Average is Difficult to Predict. *Journal of Educational Psychology* 66: 9–14.

Hamilton, L. C. 1981. Sex Differences in Self-Report Errors: A Note of Caution. *Journal of Educational Measurement* 18: 221–228.

Hunter, R. V. and Slaughter, C. D. 1980. *ETS Test Sensitivity Review Process*. Princeton, N.J.: Educational Testing Service.

Hyde, J. S. 1981. How Large are Cognitive Gender Differences? *American Psychologist* 36: 892–901.

Kappy, K. and Donlon, T. 1977. Test Speededness as a Factor in Sex Differences. Presented at the National Council on Measurement in Education Annual Meeting, New York, April.

Levitz, R. S. 1981. *The Predictability of Academic Achievement for Nontraditional- and Traditional-Age Freshmen*. Unpublished doctoral dissertation, University of Michigan.

Linn, M. C. with A. C. Petersen, M. Tobin-Richards, L. Crockett, and C. Weissbrod. 1981. Facts and Assumptions about the Nature of Sex Differences. Discussed at AERA Special Interest Group on Women in Education presession, October.

Linn, R. L. 1973. Fair Test Use in Selection. *Review of Educational Research* 43: 139–161.

———. 1982. Selection Bias: Multiple Meanings. Presidential Address to the Division of Evaluation and Measurement at the Annual Meeting of the American Psychological Association, August.

Linn, R. L. and Werts, C. E. 1971. Considerations for Studies of Test Bias. *Journal of Educational Measurement* 8(1): 1–4.

Lockheed, M. E. 1982. Sex Bias in Aptitude and Achievement Tests Used in Higher Education. In *The Undergraduate Women: Issues in Educational Equity*. Peron, P. J., ed. Lexington, Mass.: Lexington Books.

Lord, F. M. 1977. *Practical Applications of Item Characteristic Curve Theory* (Research Bulletin 77-3). Princeton, N.J.: Educational Testing Service.

Maccoby, E. E. and Jacklin, C. M. 1974. *The Psychology of Sex Differences*. Standford, Calif.: Stanford University Press.

Munday, L. A. 1976. *Declining Admissions Test Scores* (ACT Research Report No. 71). Iowa City, Ia.: The American College Testing Program.

National College-Bound Seniors, 1981. 1981. Princeton, N.J.: Admissions Testing Program of the College Board.

National College-Bound Seniors, 1983. 1983. Princeton, N.J.: Admissions Testing Program of the College Board.

Novick, M. R. 1982. Educational Testing: Inferences in Relevant Subpopulations. *Educational Researcher* 11: 4–10.

On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. 1977. New York: College Entrance Examination Board.

Pallas, A. M. and Alexander, K. L. 1983. Sex Differences in Quantitative SAT Performance: New Evidence on the Differential Coursework Hypotheses. *American Educational Research Journal* 20: 165–182.

Ramist, L. and Arbeiter, S. 1984. *Profiles, College-Bound Seniors, 1983*. New York: College Entrance Examination Board.

Sawyer, R. and Maxey, E. J. 1982. *The Relationship Between College Freshman Class Size and Other Institutional Characteristics and the Accuracy of Freshman Grade Predictions* (ACT Research Report No. 82). Iowa City, Ia.: American College Testing Program.

Strassberg-Rosenberg, B. and Donlon, T. F. 1975. Content Influences on Sex Differences in Performance on Aptitude Tests. Presented at the Annual Meeting of the National Council on Measurement in Education, Washington, D.C., March.

Tittle, C. K., McCarthy, K., and Steckler, J. F. 1974. *Women and Educational Testing*. Princeton, N.J.: Educational Testing Service.

Welch, W. W., Anderson, R. E., and Harris, L. J. 1982. The Effects of Schooling on Mathematics Achievement. *American Educational Research Journal* 19: 145–153.

Wheeler, P. and Harris, A. 1981. *Comparison of Male and Female Performance on the ATP Physics Test* (College Board Report No. 81-4). New York: College Entrance Examination Board.

Wild, C. L. 1977. Statistical Issues Raised by Title IX Requirements on Admission Procedures. *Journal of National Association of Women Deans, Administrators, and Counselors* 40: 53–56.

Wild, C. L. and Dwyer, C. A. 1977. Sex Bias in Selection Procedures and Selection Instruments. Presented at the Third International Symposium on Educational Testing, Leyden, The Netherlands.

Wittig, M. A. and Petersen, A. C., eds. 1979. *Sex-Related Differences in Cognitive Functioning*. New York: Academic Press.